

# **AUTOMATED FARMING PREDICTION**



**Inspiring Excellence**

**Talha Siddique      16241001**

**Dipro Barua      12201029**

**Zannatul Ferdous      12101045**

**Supervisor: Dr. Amitabha Chakrabarty**

**Department of Computer Science and Engineering**

**School of Engineering and Computer Science**

**BRAC University**

**Submitted On: 14<sup>th</sup> December, 2016**

# DECLARATION

Thesis Submission to the Department of Computer Science and Engineering, BRAC University, Dhaka, submitted by the authors for the purpose of obtaining the degree of Bachelor of Science in Computer Science, and the degree of Bachelor of Engineering in Computer Science and Engineering. We hereby announce that the results of this thesis are entirely based on our research. Resources taken from any research conducted by other researchers are mentioned through reference. This thesis either in whole or in part, has not been previously submitted for any degree.

**Signature of Supervisor:**

\_\_\_\_\_

Dr. Amitabha Chakrabarty

Assistant Professor

Department of Computer Science and Engineering

BRAC University

**Signature of Authors:**

1. \_\_\_\_\_

Talha Siddique

2. \_\_\_\_\_

Dipro Barua

3. \_\_\_\_\_

Zannatul Ferdous

# ABSTRACT

Farming in Bangladesh is mostly done manually. The automated way of farming here is still not introduced. This research is trying to apply a fundamental approach to inaugurate the automated process in farming in our country. It is an automated farming system designed in android application, which has been implemented to choose the best crop before starting the cultivation process according to the area of the cultivating land. Here, the best crop signifies the crop which will be the most cost effective for that particular land. In this case, the six major crops of Bangladesh – Aus, Aman, Boro, Potato, Wheat and Jute will be considered. This system is also able to prepare a schedule of total cultivation process e.g. the correct time of fertilization and irrigation according to the kind of crop types. The total system is focused on the climate and geographical condition of different areas of Bangladesh. It predicts the best cost effective crop using a prediction based algorithm. The algorithm are aimed to use is multiple linear regression with the association of some independent variables i.e. rainfall, average maximum temperature and average minimum temperature of certain location and give prediction based on yield rate per unit area. Later, KNNR algorithm was used to compare the accuracy and error rate of the predicted yield rate. To describe the functionality of this system; at first, farmer gives the perimeter of land in input area and the district from dropdown menu if he wants the suggestion of best crop. Then best crop name will be shown in the screen. If the suggestive crop is chosen, the entire steps of cultivation will be shown to him. Then the notification of irrigation, fertilization will be shown up timely or in a calendar form. The crop zone is divided according to the division and districts. The data of crops of total seven regions – Bogra, Comilla, Dinajpur, Sylhet, Dhaka, Barisal, Faridpur, Khulna, Rajshahi and Rangpur will be stored in database system. The dataset consists of information on six major crops of Bangladesh; their yield rate, maximum temperature, minimum temperature, year range, region and rainfall. The past twelve years (2000-2011) of Bangladesh have been considered making this dataset to ensure learning and training of the algorithm and increasing the accuracy rate of the prediction and for testing we used three years (2012-2014) for computing accuracy.

**Index term:**Multiple Linear Regression Analysis (MLR), prediction, KNNR, Android Application, Fertilizer Suggestion, Dependent Variable, Independent Variables.

# ACKNOWLEDGEMENT

Idea of this thesis project came to the mind of Dr. Amitabha Chakrabarty, Assistant Professor of the School of Computer Science and Engineering of BRAC University. He thought the three of us – Talha Siddique, Dipro Barua and Zannatul Ferdous worthy to carry out this thesis successfully. Each and every member of this group has contributed their level best for this thesis and we are finally presenting a good work.

We would like to express our heartiest thanks and utmost gratitude to our supervisor; Dr. Amitabha Chakrabarty, Assistant Professor of the School of Computer Science and Engineering of BRAC University for providing his valuable insight, suggestion and guidance at our thesis progress. A continuous supervision and the spirit to push us more for getting a great thesis project and an astonishing paper on this would never be possible without him. We are honored that he allowed us to work under his supervision.

We would like to remember contribution of our friends who stood beside us through the well and illness. Some friends we are grateful to in this occasion, are – AKM Nivrito, Rupa Barua, Shauvik Sadman, Hasnine Wafiq, Mohannad Zoha, Moneruzzaman Manik and many more. We acknowledge the help Arzon Barua, Web Developer of Beximco IT and Tushar Kanti Roy, Senior Software Engineer of Enosis Solutions provided us with.

Family is something that was always there for us. With the financial, tactical and mental support, they kept us on the right track throughout this journey. Even in some crucial tasks like data collection, process validation and such; we had help from our parents. Without them, we would never have been able to come this far.

We are thankful to Agriculture Information Service (AIS) for providing us with the necessary data to incorporate our research. Our project had to have a thorough analysis of the data and had to go through so many trial and error stages. We thank MD. Zakir Hasnat, Information Officer, Plant Protection of AIS for guiding us in this way. We thank him for supporting us to gather necessary data from BBS (Bangladesh Bureau of Statistics).

Lastly, we convey gratitude from the depth of our hearts towards the faculty members of Computer Science and Engineering department for providing us with all the knowledge throughout our undergraduate life so that we could do this thesis.

# Table of Contents

---

Title NamePageNo.

## **Chapter 1 - Introduction**

1.1 Introduction.....	1
1.2 Motivation.....	2
1.3 Thesis Contribution.....	3
1.4 Problem Statement.....	4
1.5 Solutions .....	4
1.6 Methodology .....	5
1.7 Summary of Contributions .....	5
1.8 Goals.....	6
1.9 Thesis Outline.....	7

## **Chapter 2– Research and Literature Review**

2.1 Supervise Machine Learning.....	8
2.2 Linear Regression.....	8
2.3 K Nearest Neighbors (KNN) Regression.....	9
2.4 Application of Linear Regression and KNN.....	10
2.5 Pseudo code for Multiple Variable Linear Regression.....	10
2.6 Pseudo code for KNN Regression.....	11

## **Chapter 3 – Algorithm and System Architecture**

3.1 Six Major Crops of Bangladesh.....	12
3.2 The Sample Area.....	12-13
3.3 Room for Improvement.....	14
3.4 System Work Flow.....	14-18
3.5 Use Case Diagram.....	19-21
3.6 Class Diagram.....	22-25

## **Chapter 4 – Dataset and Result Analysis**

4.1 Result Analysis of Barisal region.....	27-28
4.2 Result Analysis of Bogra region.....	29-30
4.3 Result Analysis of Comilla region.....	31-32
4.4 Result Analysis of Dhaka region.....	33-34
4.5 Result Analysis of Dinajpur region.....	35-36
4.6 Result Analysis of Faridpur region.....	37-38
4.7 Result Analysis of Khulna region.....	39-40
4.8 Result Analysis of Rajshahi region.....	41-42
4.9 Result Analysis of Rangurregion.....	43-44
4.10 Result Analysis of Sylhet region.....	45-46
4.11 Data Analysis before and after adding Synthetic Data, Removing Outliers.....	47-50

## **Chapter 5 - Mobile Application**

5.1 Feasibility Analysis.....	51-52
5.2 Requirement Analysis.....	52-53
5.3 User Interfaces.....	54-60
5.4 Features.....	61-62
5.5 Limitations.....	62

## **Chapter 6 - Conclusion**

6.1 Conclusion.....	63-64
6.2 Future Work.....	64

<b>References.....</b>	<b>65-66</b>
------------------------	--------------

## Table of Figures

Figure No.	Figure Name	Page
2.2	A Simple Regression Model	9
3.4a	Initial Work Flow	15
3.4b	Final Work Flow	17
3.5	Use-Case Diagram	19
3.6	Class Diagram	22
4.1a	Resulting Percentage Error of K-NN and Linear Regression for Barisal in 2012	27
4.1b	Resulting Percentage Error of K-NN and Linear Regression for Barisal in 2013	27
4.1c	Resulting Percentage Error of K-NN and Linear Regression for Barisal in 2014	28
4.2a	Resulting Percentage Error of K-NN and Linear Regression for Bogra in 2012	29
4.2b	Resulting Percentage Error of K-NN and Linear Regression for Bogra in 2013	29
4.2c	Resulting Percentage Error of K-NN and Linear Regression for Bogra in 2014	30

4.3a	Resulting Percentage Error of K-NN and Linear Regression for Comilla in 2012	31
4.3b	Resulting Percentage Error of K-NN and Linear Regression for Comilla in 2013	31
4.3c	Resulting Percentage Error of K-NN and Linear Regression for Comilla in 2014	32
4.4a	Resulting Percentage Error of K-NN and Linear Regression for Dhaka in 2012	33
4.4b	Resulting Percentage Error of K-NN and Linear Regression for Dhaka in 2013	33
4.4c	Resulting Percentage Error of K-NN and Linear Regression for Dhaka in 2014	34
4.5a	Resulting Percentage Error of K-NN and Linear Regression for Dinajpur in 2012	35
4.5b	Resulting Percentage Error of K-NN and Linear Regression for Dinajpur in 2013	35
4.5c	Resulting Percentage Error of K-NN and Linear Regression for Dinajpur in 2014	36
4.6a	Resulting Percentage Error of K-NN and Linear Regression for Faridpur in 2012	37
4.6b	Resulting Percentage Error of K-NN and Linear Regression for Faridpur in 2013	37



4.6c	Resulting Percentage Error of K-NN and Linear Regression for Faridpur in 2014	38
4.7a	Resulting Percentage Error of K-NN and Linear Regression for Khulna in 2012	39
4.7b	Resulting Percentage Error of K-NN and Linear Regression for Khulna in 2013	39
4.7c	Resulting Percentage Error of K-NN and Linear Regression for Khulna in 2014	40
4.8a	Resulting Percentage Error of K-NN and Linear Regression for Rajshahi in 2012	41
4.8b	Resulting Percentage Error of K-NN and Linear Regression for Rajshahi in 2013	41
4.8c	Resulting Percentage Error of K-NN and Linear Regression for Rajshahi in 2014	42
4.9a	Resulting Percentage Error of K-NN and Linear Regression for Rangpur in 2012	43
4.9b	Resulting Percentage Error of K-NN and Linear Regression for Rangpur in 2013	43
4.9c	Resulting Percentage Error of K-NN and Linear Regression for Rangpur in 2014	44
4.10a	Resulting Percentage Error of K-NN and Linear Regression for Sylhet in 2012	45

4.10b	Resulting Percentage Error of K-NN and Linear Regression for Sylhet in 2013	45
4.10c	Resulting Percentage Error of K-NN and Linear Regression for Sylhet in 2014	46
4.11a-d	Data Analysis for Aus of Dhaka region	47
4.11e-h	Data Analysis for Aman of Faridpur region	48
4.11i-l	Data Analysis for Boro of Khulna region	49
5.3a	Log In/ Register	54
5.3b	Welcome UI	54
5.3c	Prediction	55
5.3d	Choosing Crop	55
5.3e	Cropping Procedure	56
5.3f	Fertilizer Suggestion	56
5.3g	Setting Notification	57
5.3h	Reminder of Harvesting	57
5.3i	Side Button	58
5.3j	Feedback	58

5.3k	About Section	58
5.3l	Invalid Username	59
5.3m	Invalid Password	59
5.3n	Data Choosing	60
5.3o	Wrong Date	60

# Chapter 1

## INTRODUCTION

In Introduction, we will describe about our system briefly and give a complete overview of the system. The functionality and features of the Automated farming prediction farming app are discussed in this chapter. Later other part of this report, contain more information about this app.

### 1.1 Introducton

Automated farming is a predictive android application which is able to predict the most cost effective crop by taking input of location and planting date by farmers. The machine learning algorithm [2][3] which has been incorporated in this android application is Multiple Liner Regression [6][3]. Later on, we have shown a comparative study with K Nearest Neighbor Regression (KNNR) [7] to show the accuracy and error rate of the predictive outcomes. We are dividing the crop zones according to the division and districts. There is information about the crops of total seven divisions namely - Chittagong, Sylhet, Dhaka, Barisal, Khulna, Rajshahi and Rangpur. The dataset consists of information on "6 major crops of Bangladesh" namely - Aman Rice, Aus Rice, Boro Rice, Potato, Wheat and Jute. In this system, farmer will give input the date of the plantation and the region wise location. Then at the back end of the crop, MLR[6] runs and on the screen the predicted crop along with the yield rate appears chronologically. It will perform prediction in terms of percentage change in output per unit area based on previous records. The independent variables are average temperature and rainfall. It also increases the farmer's profit margin by giving the appropriate suggestion of crops. Our goal is to make a project prototype that will be easy to operate even for amateurs in technology usage. To make sure that a cost-effective farming solution is given to the farmer, taking cultivation location [1] (zone/district) as the input; our app suggests the most effective crop type to be cultivated. Agriculture being one of the most important sectors of income in this country [4] had far less involvement of technology. The mobile application will help farmers to have access to an easily operating application with simple interface and more updated database. This app will help a farmer to remember different farming schedules (water, pesticide, fertilizer, irrigation) through scheduling notifications. To keep in mind the education level of most of the farmers in our country, this app will be available in Bangla.

We are focusing on more offline usage of this android application rather than its online use. The communication system of rural area of our country is not advanced as our urban area. There are no Wi-Fi or high speed internet connection. So, we are considering that fact and whenever it contacts with the internet, the database will be updated and the new data will be stored to the mobile's local database to use this app offline. This app's GUI interface will be user friendly and manageable by our farmers.

## **1.2 Motivation**

The main motivation behind this research was using the information technology in farming sector first time in our country. Building this automated system for farming was not very easy for us, we had to collect data from the scratch and by using this data bring the appropriate output prediction was very tough job. To make this application successful for the sake of our country was main motivation for us.

Our goal is to build a farming app that can support the Asian, to be precise the Bangladeshi farmers. The developed countries of the world, cultivate crops in a very advance way [5] but in our country, the crop cultivation is a long manual process. There is less application of modern technology and method in this sector. This system will ensure to farm in a technological way. It will ensure proper crop allocation throughout the country because three fifths of Bangladeshis are employed [4] in the agriculture sector, so this will increase our total production of agriculture and maximizes crop allocation throughout the country. This will also impact in GDP growth [4] in our economic system. The future generations will also want to involve in agricultural sector.

In our thesis, we propose an android application which will use MLR [6] to predict the location wise best crop among six major crops of Bangladesh and generate a total farming schedule which will as well as indicate the amount of resources are going to need in the farming field.

### **1.3 Thesis Contribution**

This thesis contributes to the society with the application of android technology and machine learning algorithm [2][3] for predictive measures in our agricultural sector. Prior to this work, no work like this was performed in our country. We have done the proper market analysis and made feasibility study. The technological feasibility, economical and organizational feasibility was determined before starting of this research. Then we have done the requirement analysis on the basis of functional and non-functional requirement. The functional requirement considered the availability of smart phone in rural area, internet connection, and user authentication and so on. Then, the non- functional requirement considers the UI that support the ease of use of the farmers, quick recovery of the system after failure and performance of the system. This total system was fully organized from the beginning considering all the facts and pros and cons about this mobile app. By doing this kind of analysis we were able to find loopholes of this system from the very beginning and were able to fix it step by step. This android application is now mostly perfect to perform any kind of action from its specification. Now, we know there is almost no system flaws with this system those we often find in other systems and its proper market values. The benefit of the existence of this app and its importance in the agricultural sectors was identified. We feel proud that, this is the first automated approach that is performed in our agriculture-based sector.

## **1.4 Problem Statement**

The main impediment of this thesis was data collection from the handwritten copies. Then we had to store these data into various format to perform the train and test on them with the machine learning algorithm. The year wise data collection was not easy. There was lacking of digital approach to save the agricultural data in various government organizations as well. We had to overcome those difficulties. Then the next hindrance was working with bulk amount of data and there were many extreme situations as well. To make our prediction successful, we had to fix our code and procedures many times. We were able to make a prototype android application after dealing with these problems. Another hindrance was the limited time constrains to complete this project. We had to fight with our packed schedule with our classes and examinations to work in this thesis. A lot of dedication was needed to complete and reach this research up to this point. Another big problem was lacking of knowledge about the correct cultivation process. We had to study a lot and talk with the AIS a lot to find the perfect procedure to cultivate each crops. Other problem of this research was the knowledge gap. We had to work with the machine learning algorithm, in which we had no prior mastery or knowledge. We had to start it from the scratch to understand and apply these algorithms in our dataset. There was limitation of investing money for the larger database and its performance issue. For the time constrains, we couldn't verify our result with the actual outcome from the farming field. That will be the measuring scale of our system if this application can give the perfect result with less error percentage. We couldn't incorporate the sensors to know the composites of soil of our very limited time constrains. So, the more dynamic features can be added in later.

## **1.5 Solutions**

First of all, using the MLR [6] algorithm was the best choice to give prediction using this dataset. Then by omitting some outlier from our dataset we got better accuracy with less error percentage in prediction. So, those were the solution for our system. After that, we have done some comparative study with KNNR [7], in the six regions to show that, MLR [6] gives the best accuracy over KNNR [7].

## **1.6 Methodology**

In the literature review, we will see that the MLR [6] and KNNR [7] algorithms are used by so many researchers about crop prediction in different countries. We are also using these algorithms in our research to give the appropriate predictions. In our android application, we used Multiple Linear Regression in the prediction portion. From the database, data is pulled and these algorithms run in the application. Later, the results of yield per unit area of different crops have been compared and the best two crops according to the best yield rate are shown in the app screen. After choosing the crop, either our predicted crops or by his own choice, the total schedule of farming along with appropriate resources will be shown. By the time being, the farmer will get the notification of applying irrigation, pesticide and fertilizer in his smart phone. If he wishes, he can do the task within the five days after getting notified. There are some features of this application that will be supported by the earlier discussion has been added into the system. They are login by one click process to reduce hassle, the freedom of choosing crops by farmers. For example, if the app suggests the crop will most cost effective is potato. On the contrary, farmer wanted to cultivate rice in this land. He can choose according to his will and as well as get the notifications and all procedures about that crop. The servers of data were managed efficiently. Such as, local server for fast processing and main server to contain all necessary data. All the described features will be available in bundle for better understanding of the farmers.

## **1.7 Summary of Contributions**

The work presented in this thesis contributes to improving the quality of farming in our country and the algorithms used to the application and comparison are already present in the literature review. Automated way of farming is the future of the farming. We have started the attempt by making a prototype but there will be huge space of work for the future. Along with the crop prediction, the flood prediction is equally important in farming sector. So, later on, this research will act as a primary footstep in this field. By doing further research on flood, using sensor for soil composition will help a lot of farmers in our country.

In this thesis, we have shown that it is indeed possible to predict appropriate crop for farming within respective regions. With the help of the farming information, it is also possible to generate an automated schedule and set the schedule notifications.



## 1.8 Goals

In our research, our main objective is to come up with an algorithm that gives better prediction about appropriate crop according to the region and crop season in our country. Here we used the MLR [6] to serve the purpose using rainfall and average temperature as dependent variables and yield rate per unit area as independent variable. We have succeeded to get appropriate prediction values by using this algorithm. Another objective was, to introduce automated farming process in our country. Introduce technology in our rural area and make our farming sector more technologically advanced. By implementing our prototype after some more research, this system can greatly serve our country. Another goal is to make farmer's profit margin high and increase the GDP growth in our country [4]. If we can successfully implement machine learning algorithm [6][7] in our agricultural sector, later on, it will be able to predict and implement more useful things in the production of crop. Such as, useful prediction for flood, natural calamities using weather update, dynamic fertilizer prediction and so on. So, this work will open more possibilities in the sector of crop farming using machine learning technology. It will promote the future aspects of our agriculture and farming policy and method. Another goal is, to make familiar to our agricultural process to the outside world by the mobile application process. If we become successful, then other developing countries will be encouraged to use our system or take our idea to make a system similar to this. This will make our country more familiar to the outside world and bring fame and praise for us. With the sufficient help from our government this goal can be achieved and can be executed properly.

## **1.9 Thesis Outline**

Chapter 1 gives a brief overview of our research, our estimated goal and what we have gained.

Chapter 2 discusses the literature review and background study of our project, what algorithm were considered, how they work and more.

Chapter 3 discusses about the chosen algorithm and the reason behind choosing it. Also discusses about the system architecture, describes its functionality through diagrams.

Chapter 4 discusses comparative study and analysis of the resulting output we found. The trial and error stages are also briefly discussed in this chapter.

Chapter 5 focuses on the mobile application. Functions and limitations of the app, screenshots discussing real life implementations, requirements and feasibility analysis are also part of this chapter.

Chapter 6 ends our paper with conclusion and proposed future works for the system.

## **Chapter 2**

### **LITERATURE REVIEW**

This chapter gives an overview and explanation of machine learning algorithms: linear regression and K Nearest Neighbor. It also discusses some of the relevant works that has been done using these algorithms in other sectors and how these works motivated us to implement these algorithms in the agricultural farming sector.

#### **2.1 Supervised Machine Learning**

Supervised machine learning algorithms consist of a target/ outcome variable (or dependent variable) which is to be predicted [3] from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data.

#### **2.2 Linear Regression**

It is a form of supervised machine learning [2] which is used to estimate real values based on continuous variable(s). Here, we establish relationship between independent and dependent variables by fitting a best line. This best fit line is known as regression line and represented by a linear equation  $Y = a * X + b$ .

In this equation:

- Y – Dependent Variable
- a – Slope
- X – Independent variable
- b – Intercept

These coefficients “a” and “b” are derived based on minimizing the sum of squared difference of distance between data points and regression line.

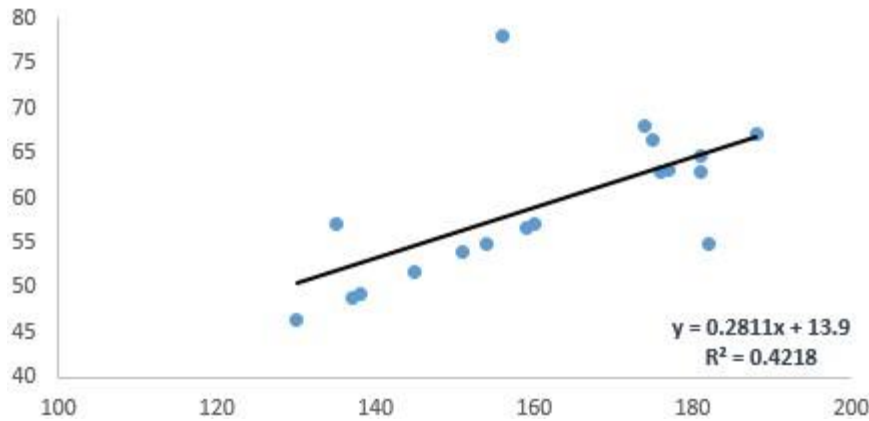


Figure 2.2: A Simple Regression model [3]

Linear Regression is of mainly two types [3]: Simple Linear Regression and Multiple Linear Regression. Simple Linear Regression is characterized by one independent variable. And, Multiple Linear Regression is characterized by multiple independent variables. While finding best fit line, you can fit a polynomial or curvilinear regression. And these are known as polynomial or curvilinear regression.

### 2.3 K Nearest Neighbors (KNN) Regression

K nearest neighbors is a simple algorithm that stores all available cases and predicts the numerical target based on a similarity measure (e.g., Euclidean distance). The optimal value for K is chosen based on the data set. In general, a large K value is more precise as it reduces the overall noise.

Euclidean Distance Formula:

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Where, x and y are the two predictors in the model.

## 2.4 Application of Linear Regression and KNN

Linear regression has been extensively used in the prediction of factors related to the stock market [8]. Studies have been carried out to predict stock market trading volumes [8]. It has also been applied to predict stock market prices [9]. Research works have been conducted using KNN to predict gene function [10]. It has been used to predict the growing stock of inventory systems [11]. This was the motivating factor behind apply both multivariable linear regression and KNN to observe which algorithm gives a better prediction of output per unit area of six major crops of Bangladesh based on rainfall and temperature as the predictors.

## 2.5 Pseudo code for Multiple Variable Linear Regression

```
Regression(double [] [] x, double [] y)

Matrix x ← new Matrix(x)

// create matrix from vector

Matrix y ← new Matrix(y,N)

//find the least square solution

QRDecompositionqr ← new QRDecomposition(x)

beta ← qrsolve(y)

// Mean of y[] values

For i=0 to N

    Sum += y[i]

    Mean ←  $\frac{sum}{N}$ 

//Total variation to be accounted for

For i=0 to N

    Dev ← y[i]-mean

    Sst+= dev*dev

// variation accounted for

Matrix residuals ← x.times(beta) . minus(y)

SSE ← x.times(beta) . minus(y)
```

## 2.6 Pseudo code for KNN Regression

```
// calculating the Euclidean Distance

For i=0 to Yield .length
D[i]= distance(rainfall, temp, rainfall[i], temp[i] )
End For

//Finding the indices of 'k' no. mean Euclidean Distances
FindMinIndex(k, d, minIndex)

//calculating the predicted yield
Yield ← calculateAverage(y,minIndex)

//Calculate percentage if error
error ← percentageErrors(y1,yield)
```

## **Chapter 3**

### **ALGORITHM AND SYSTEM ARCHITECTURE**

This chapter describes which algorithm was observed to be fit for our project and what role it plays in our android application. Also, we will discuss briefly about the system specifications, take help from some diagrams to understand how the system works and have a look at the system from the inside. We will see what was our idea at the beginning about our app and its' backend support. Why we changed this idea and what was the new working functionalities of the system. We will take a look at different use-cases in the use-case diagram and see the programming view through the class diagram.

#### **3.1 Six Major Crops of Bangladesh**

According to Bangladesh Bureau of Statistics [1] the six major crops that are cultivated in Bangladesh are classified as follows:

1. Aus Rice,
2. Aman Rice,
3. Boro Rice,
4. Potato,
5. Wheat and
6. Jute.

#### **3.2 The Sample Area**

The regions we considered for our thesis according to Bangladesh Bureau of Statistics [1] are:

1. Barisal,
2. Bogra,
3. Comilla,
4. Dhaka,
5. Dinajpur,

6. Faridpur,
7. Khulna,
8. Rajshahi,
9. Rangpur and
10. Sylhet.

The dataset obtained consists of the country's divisions and under each division there are a number of districts. For each district there are yield per unit area for each of the 6 major crops produced over a year. The data set was divided into two sets: learning sets and testing sets. For each district, the yield per unit area for a particular year will be calculated for each crop. Using the learning data set, an analysis was carried out using both multivariable linear regression and KNN algorithm. The prediction by each model was compared with the yield per unit area of the test set. The percentage error between the predicted values and the actual values were calculated to determine the accuracy of each model. In this particular project's case, multivariable linear regression outperformed KNN in terms of accuracy. Both the models took into account rainfall and temperature measurements as the two predictors.

An android application was developed in which this multiple variable linear regression algorithm was incorporated. The user will give their land location as input along with the date on which the user plans to plant the crop. The program will determine the farming season under which the selected date falls. Then it will determine which of the six major crops can be planted during that season or time frame. The program will then take into account the district from the land location that was provided and predict the yield per unit area of the crops for that particular year and district. The application will then list the crops and their respective yield per unit area in descending order. The user will also have the option to choose either the application's suggestion or any of the other crops available during that particular season. Based on the user's choice the program will suggest the necessary resources and create a schedule for the entire farming process. The user will also receive notifications periodically about the schedule.



### **3.3 Room for Improvement**

The prediction model can be improved further if more number of predictors can be incorporated. For example, soil composition data can be incorporated as a predictor. However, till date, soil composition data of previous years of agricultural lands have not been officially recorded by Bangladesh Agricultural Extension and therefore we were discouraged by them to incorporate soil composition into our model. This is because whatever data is currently available is not deemed reliable by them.

### **3.4 System Work Flow**

Our plan while doing pre-thesis 2, was to have 2 ways for suggestion part – 1) Pre-Suggestion (User interested only in information and/or will do cropping in future) and 2) Post Suggestion (User already started cropping and needs a guideline from the go)! The pre-suggestion part was for users who were just interested in information, was not planned fully. The pre-suggestion part for user who will surely do cropping in the future and the post suggestion part for the users get in the similar path after a few steps like – a. Taking the location input and b. Taking the date input.

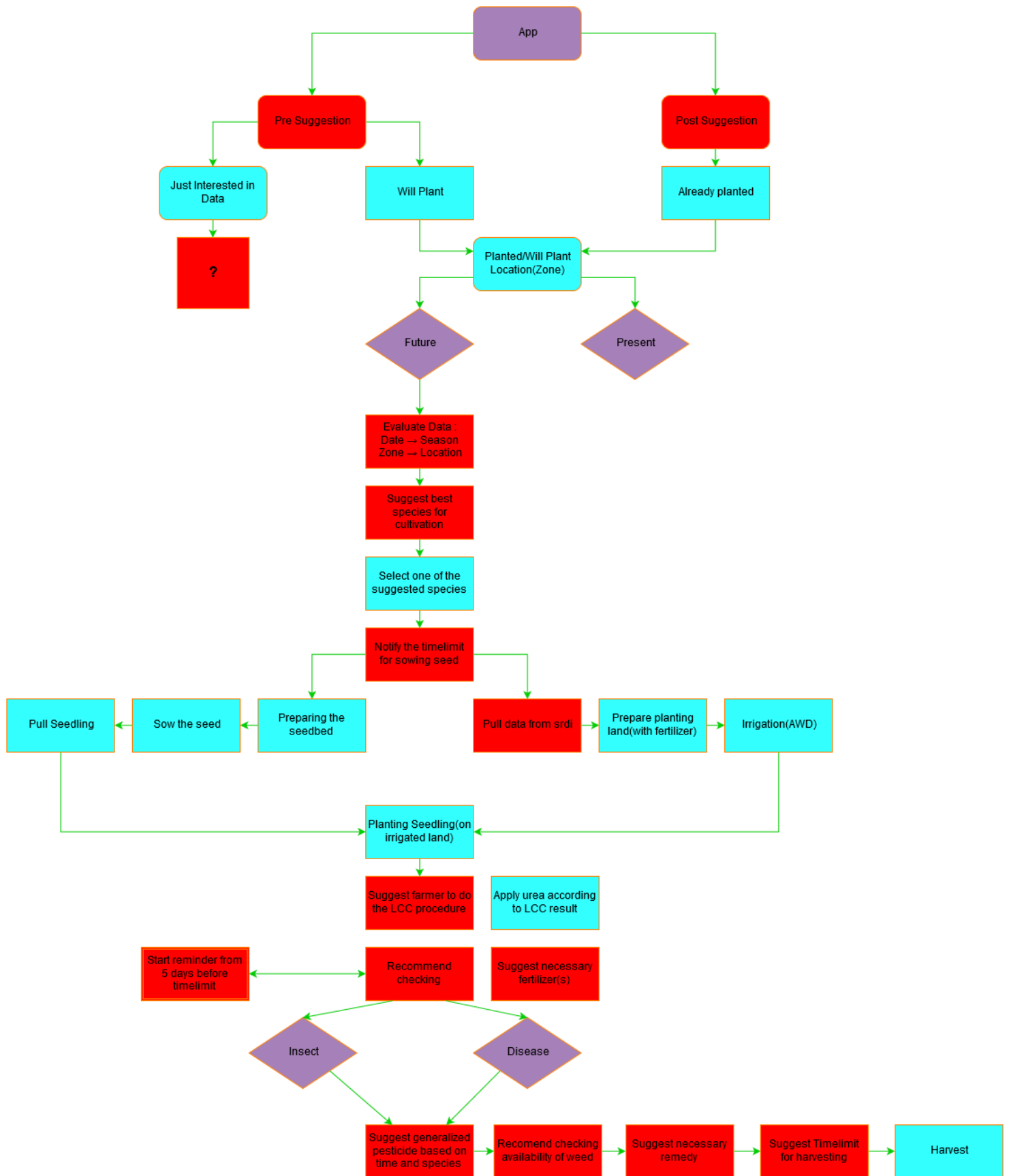


Figure 3.4a: Initial Work-Flow

And then the app would:

1. Evaluate the input – date and location,
2. Suggest the best species for cultivation,
3. Let the user select one of the suggested species,
4. Notify user the time limit for sowing seed and then app shows some different things along with this:
  - i) Time of preparing seedbed, sowing the seed, plucking the seedling,
  - ii) Pulls data from srdi for other tasks,
  - iii) Time for preparing planting land, irrigation and many more.
5. Notify the user of the time for planting seedling,
6. Suggest user to apply LCC for irrigation procedure,
7. Suggest user the necessary fertilizers,
8. Start reminder for every important task 5 days before the task,
9. Recommend user to check for disease and insects,
10. Suggest generalized pesticide and medicine and
11. Suggest time limit for the harvesting part.

After we showed this flow to the agriculture experts, they pointed out some flaws in our plan, like – a. The whole 2-way output generation is a bad start. Besides, people interested in data can always browse agricultural websites, b. Pulling data from srdi and then showing them would slow the application and c. Recommending one generalized suggestion for variable like insect and disease is not a feasible idea.

So, keeping these things in mind, we had to change our idea.

We came up with another plan and this time, we fixed our idea:

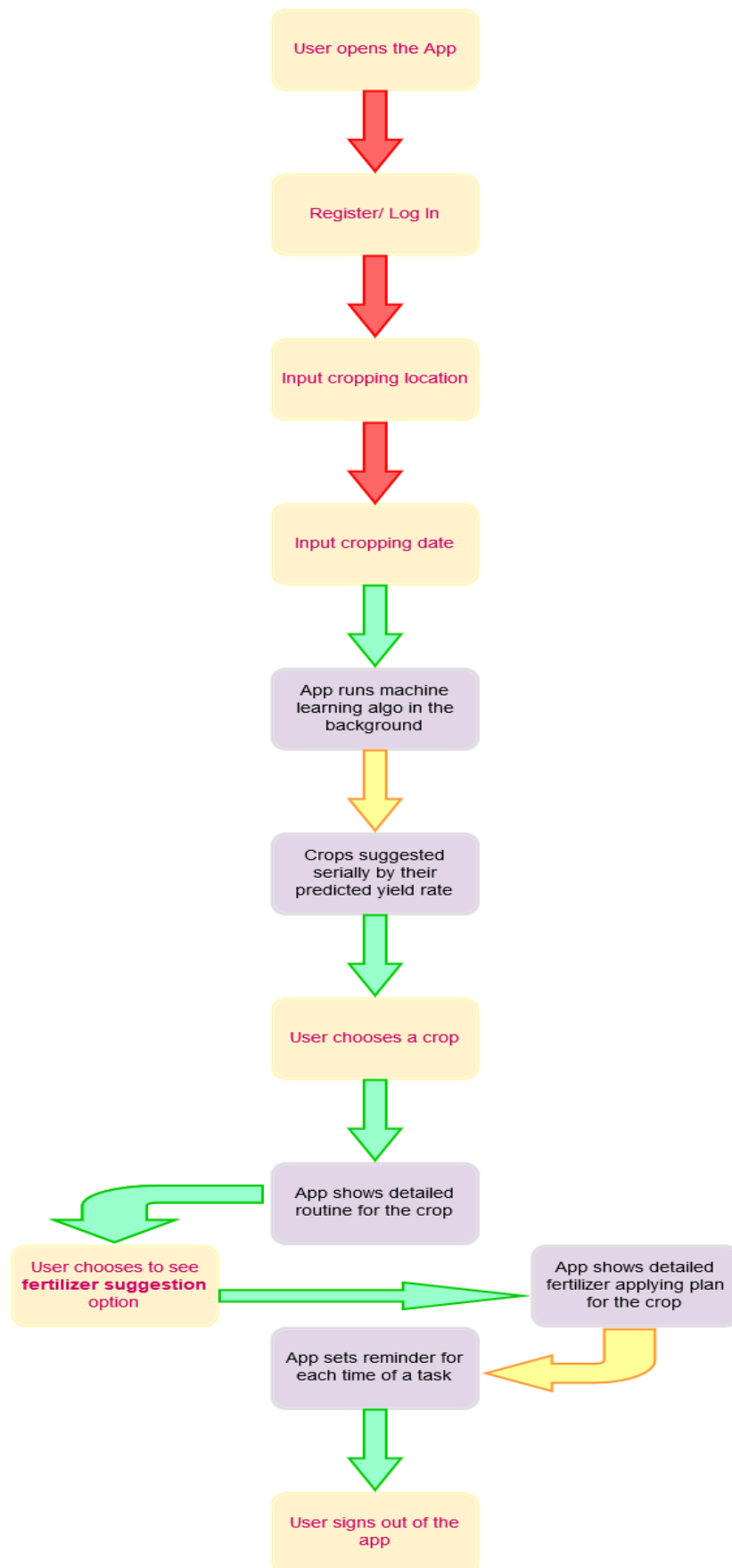


Figure 3.4b: Final Work Flow

In the finalized work flow, we removed phases of suggestion and made it only one suggestion type for all users. Current workflow:

1. User will open the app,
2. User will register/ log in to the app,
3. User will input cropping date and location,
4. App will go to the backend, using the input location and date; it will have suggested crop names,
5. Then the backend algorithm will be in work and judging data from years (2000 - 2011), it will predict yield rate for the suggested crops,
6. App will then show the suggested crops with predicted yield rates,
7. User will now choose one of the suggested crops (if there's multiple choices),
8. App will then show a detailed schedule of whole procedure throughout the cropping season for the crop,
9. User can click "Fertilizer Suggestion" button to see suggestion of fertilizer for any crop,
10. App will show tabular format as to when and how much fertilizer to be applied,
11. App then will set reminders for the future, starting from and relating to the sowing seed date,
12. User will sign out from the app.

However, after this; the app will give reminder to user on important days like – Sowing seed day, 1st irrigation day etc.

### 3.5 Use Case Diagram

Use Case representation for our project looks like:

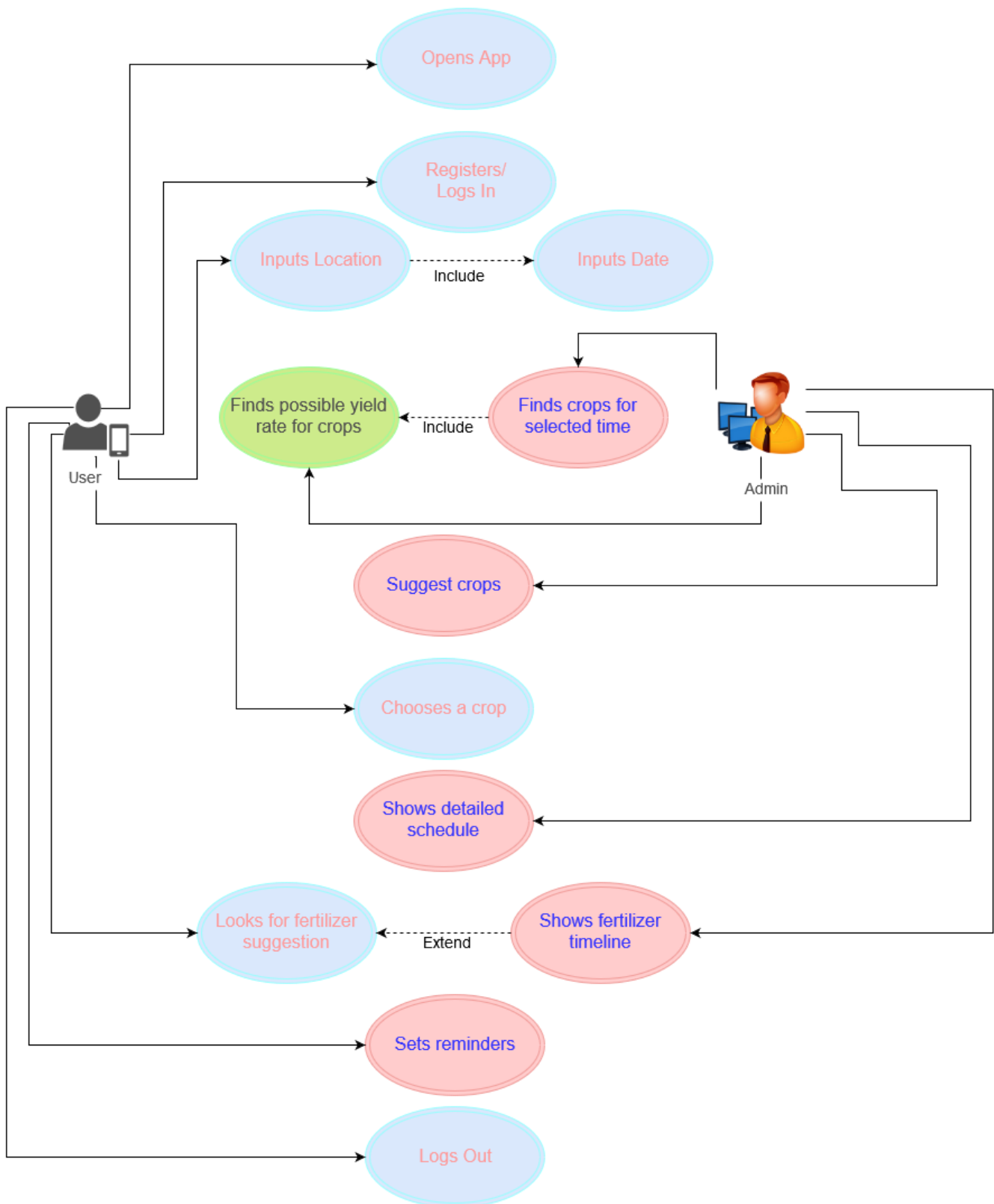


Figure 3.5: Use Case Diagram

Use Case Name: Automated Farming Prediction

Primary Actor: User

Secondary Actor: Administrator

Brief Description: The use case describes how a user operates the app, gives input to it and how the admin through the app, with the help of backend algorithm gives output to the user.

Relationships:

- Include: “Inputs Location” includes Inputs Date  
“Finds Crop for Selected Time” includes “Finds possible yield rate for crops”
- Extend: “Shows Fertilizer Suggestion” extends “Looks for Fertilizer Suggestion”

Entry Condition: The user has given a valid mobile number, a password and has registered to this app.

Exit Condition: User has knowledge of the whole cropping procedure and fertilizer distribution for the chosen crop and set notifications for the important dates.

Normal Flow of Events:

1. User opens the app,
2. User registers/ logs in to his/ her account in this app,
3. User inputs date and location for cropping,
4. App then finds out which crops are available for the chosen region and date,
5. Backend algorithm gives probable yield rate for the crops,
6. App then shows the resulted crops with their predicted yield rates, in descending order,
7. User chooses one of the crops,
8. App then shows procedure of cropping throughout cropping season,  
If the user wants to check fertilizer suggestion, then  
S-1: App shows a detailed fertilizer applying schedule
9. User exits the app  
When user is leaving the app without setting notifications, then  
S-2: App sets us reminders for important days

#### Sub Flows:

##### S-1: Fertilizer Suggestion

1. User clicks “Fertilizer Suggestion” button,
2. App shows a detailed fertilizer suggestion as to when, how much fertilizer should be applied.

##### S-2: Setting Reminders

1. User is logging out of the system.,
2. App selects user’s chosen date as “Sowing Seed” day and therefor first reminder,
3. App then sets reminder for other reminders relating to the “Seed Sowing day”.

#### External Flow:

App will give these reminders time to time to the user.

#### Exceptional Cases:

1. Internet connection is lost while user is using the app,
2. Database connection is lost for some reason,
3. Someone hacks admin id and logs in.



### 3.6 Class Diagram

A class Diagram representation of our system looks like:

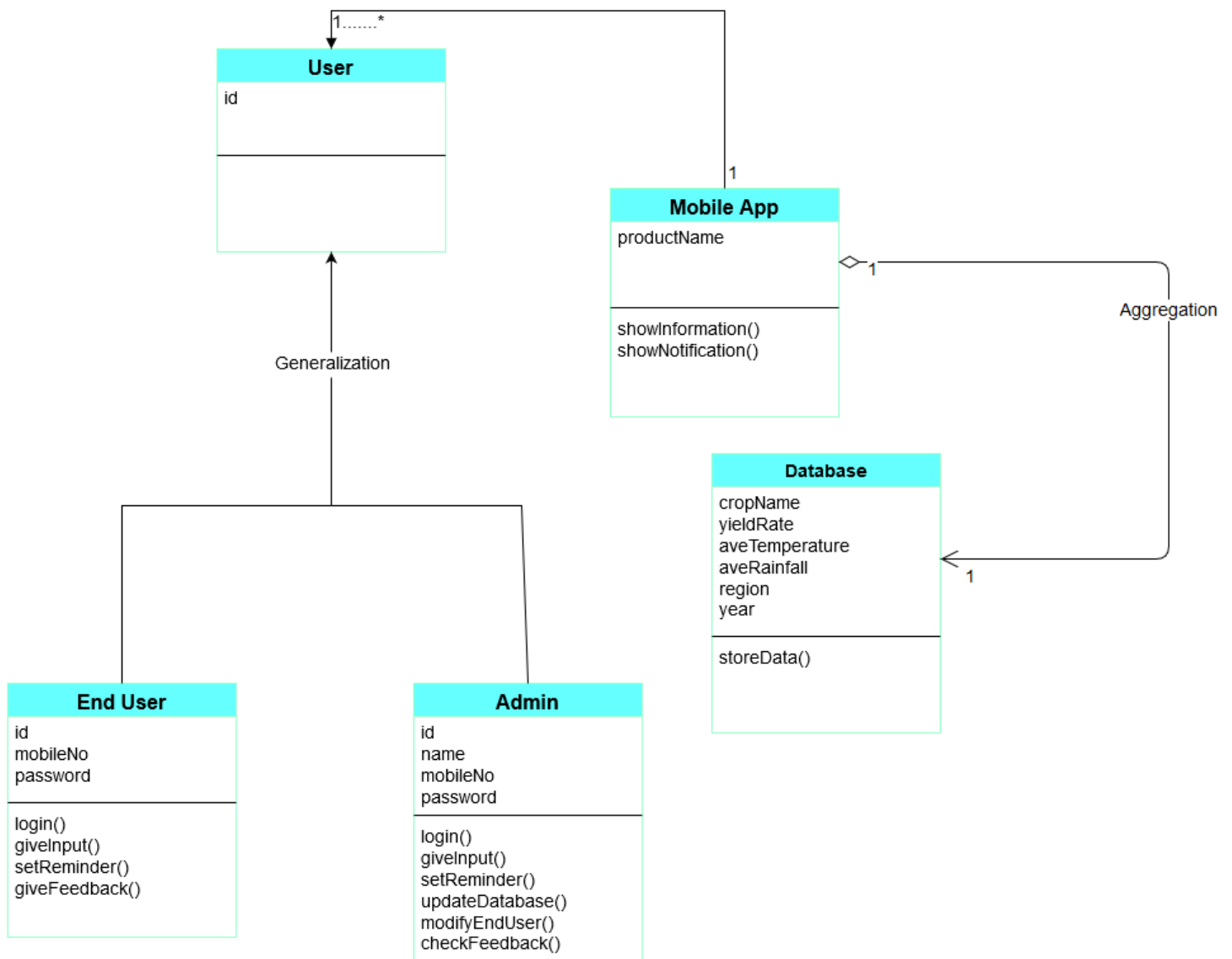


Figure 3.6: Class Diagram

The class diagram describes different attributes and methods of different classes. The classes identified are:

- a. User,
- b. Mobile App and
- c. Database.

User class has 2 specializations – 1. End User and 2. Admin.

#### A. User Class:

This class includes all the users in general. It has attribute:

- a. id

2 different sub-classes of this class are:

##### i) End User:

The sub-class includes all the users that create an account to run this app and gets benefit from it.

Attributes:

- 1) id – The number that describes a generalized serial of the users,
- 2) mobileNo – The username that users will use while creating a new account and logging in to the app,
- 3) password – The password a user will create for his/her account.

Methods:

- 1) login() – The function gets the user logging in to the app,
- 2) giveInput() – The function enables the user to give input date and location,
- 3) setReminder() – The function enables user to set notifications for important dates,
- 4) giveFeedback() – The function enables the user to give feedback on anything related to the app, its prediction and relevant things.

##### ii) Admin:

The sub-class includes all the administrative body people who have access to the database and can update any information anytime.

Attributes:

- 1) id – The number that describes a generalized serial of the admins,
- 2) name – A unique key for the admins that will distinguish them from the end users,
- 3) mobileNo – The username that admins will use while creating a new account and logging in to the app,
- 4) password – The password and admin will create for his/her account.

**Methods:**

- 1) login() – The function gets the user logging in to the app,
- 2) giveInput() – The function enables the user to give input date and location,
- 3) setReminder() – The function enables user to set notifications for important dates,
- 4) updateDatabase() – The function which allows the admins to make any modifications to the database,
- 5) modifyEndUser() – If an admin feels it necessary, s/he can modify any information and also the account itself, of any end user through this function,
- 6) checkFeedback() – Admins can check what feedbacks are stored at the database, through this function.

**B. Mobile App:**

This class mainly portrays the mobile application, it's front-end.

**Attributes:**

- 1) productName – The name of the mobile and its model.

**Methods:**

- 1) showInformation() – The function that enables the mobile application to show information,
- 2) ShowNotification() – The function that enables the application to show timely notifications to the user.

**C. Database:**

This class symbolizes the backend data support system of our application.

**Attributes:**

- 1) cropName – Name of different crop types,
- 2) yieldRate – Yield rate per year for different regions of different crop types,
- 3) aveRainfall – Average rainfall of the cropping time-period months for any chosen crop,
- 4) aveTemperature – Average temperature of the cropping time-period months for any chosen crop,
- 5) region – The chosen 10 regions,
- 6) year – The learning data and testing data, total of 15 years.

Methods:

- 1) storeData() – The function that stores all the data into the data support system.

## **Chapter 4**

### **RESULT ANALYSIS**

We have checked both Multi-Variable Linear Regression and KNN (K-Nearest neighbor) algorithm in order to determine which algorithm works better. Our workspace was the mentioned 10 regions and we chose the 6 major crops as our options.

The 3 variables for each crop we considered were – i. Crop yield, ii. Average Temperature per month and iii. Average rainfall throughout the cropping period per year.

The results we found are shown in the graphs in the next page:

#### 4.1 Result Analysis for Barisal region

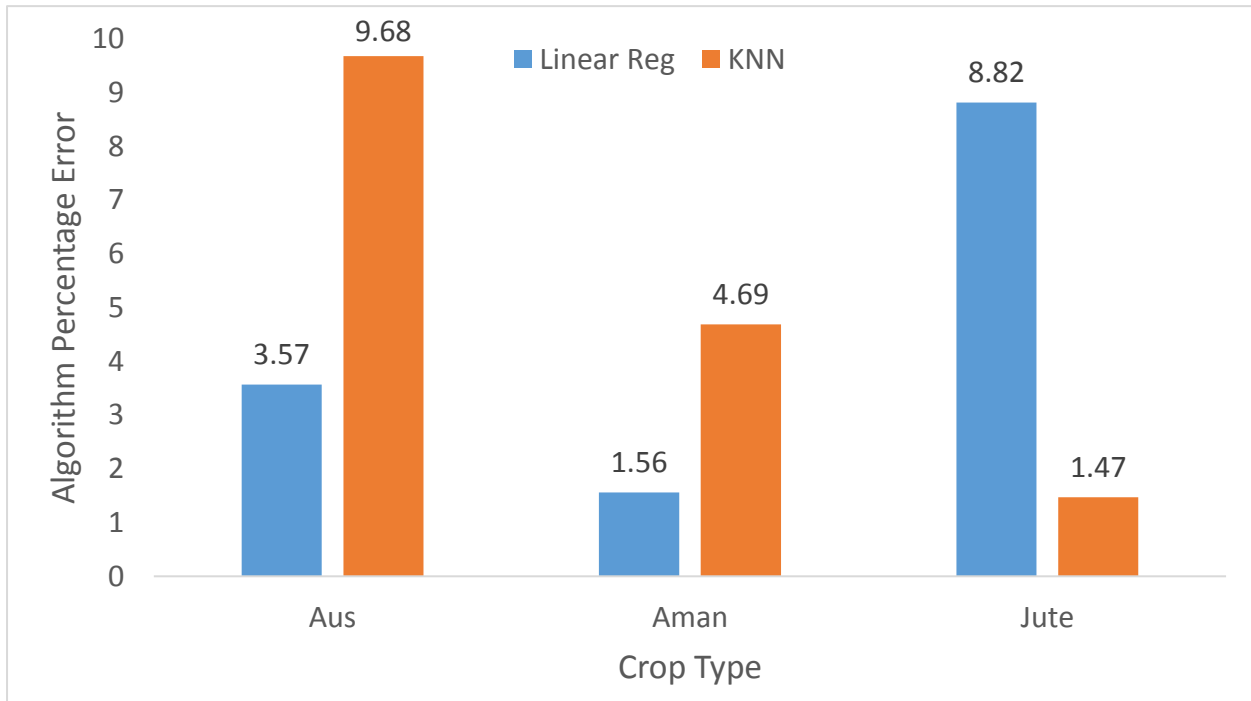


Figure 4.1a: Resulting Percentage Error of K-NN and Linear Regression for Barisal in 2012

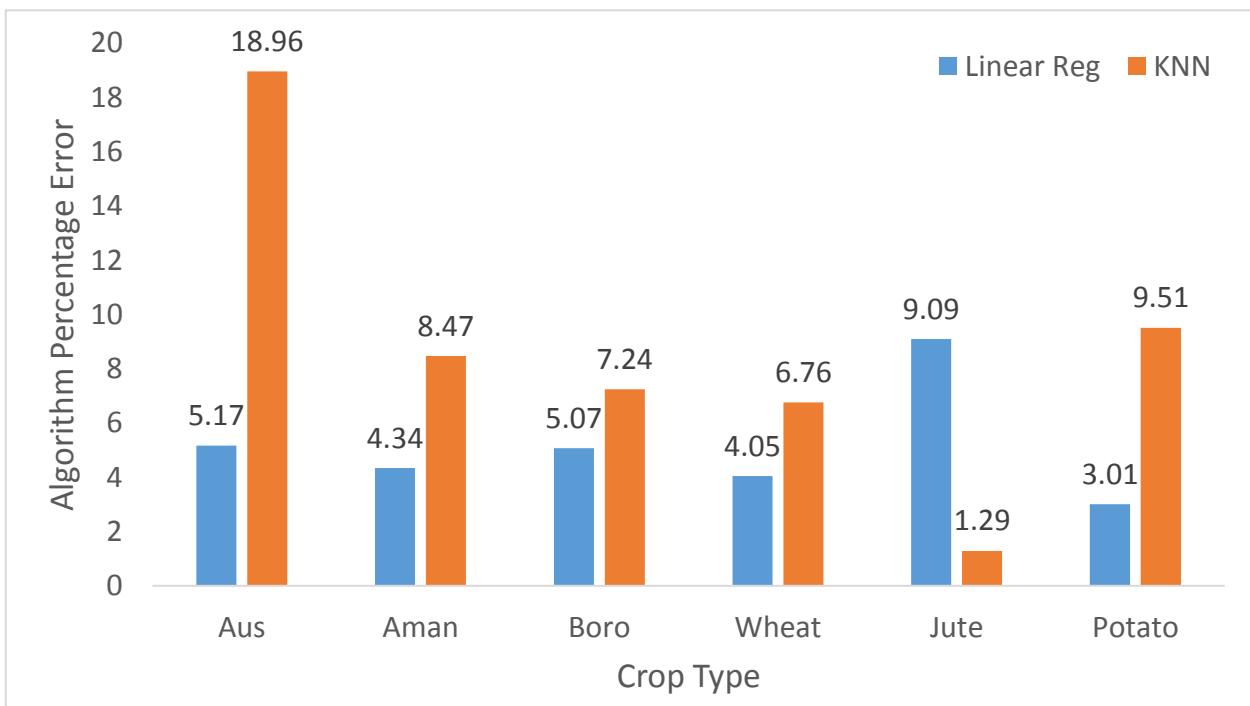


Figure 4.1b: Resulting Percentage Error of K-NN and Linear Regression for Barisal in 2013

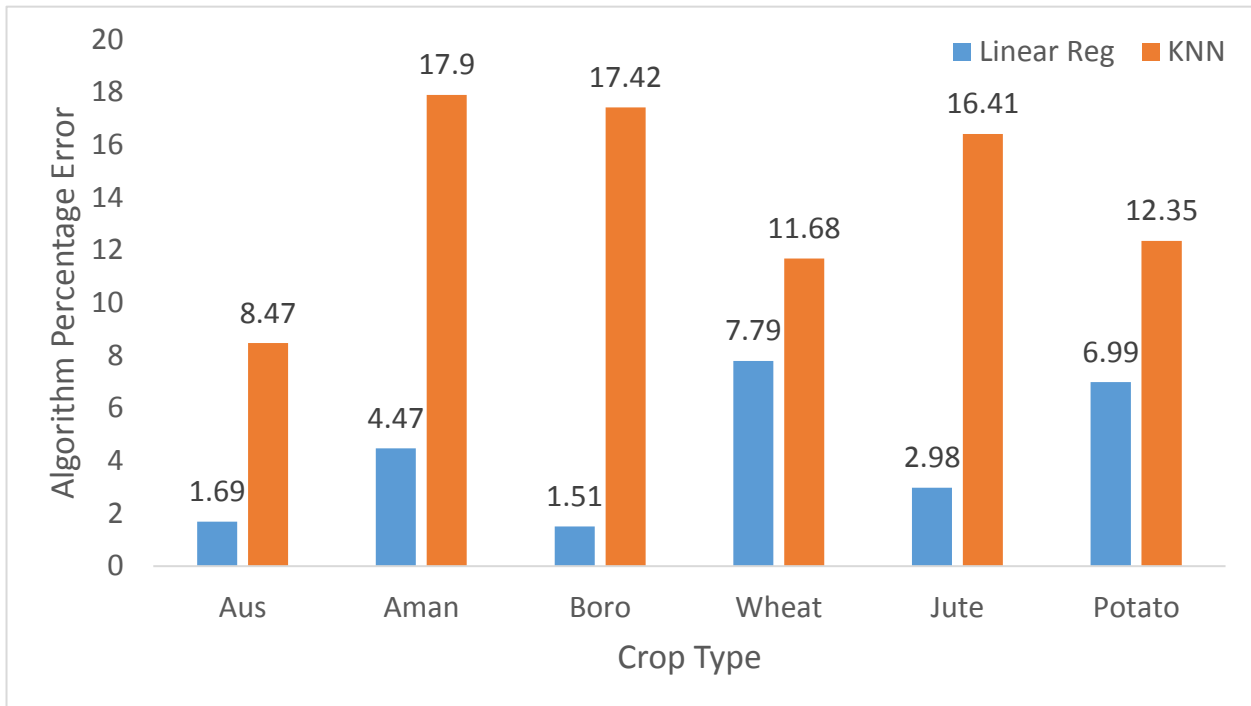


Figure 4.1c: Resulting Percentage Error of K-NN and Linear Regression for Barisal in 2014

The result analysis dataset of Barisal for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of Aus and Aman better than KNN algorithm. And KNN algorithm predicts yield rate of Jute better than Linear Regression algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of Aus, Aman, Boro, Wheat and Potato better than KNN algorithm. However, KNN algorithm predicts yield rate of Jute better than Linear Regression algorithm once again.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm.

Reason behind this type of result in the years 2012 and 2013, is that the cropping time-period of Jute (month of April to month of July) in these years had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned years for the mentioned crop.

## 4.2 Result Analysis for Bogra Region

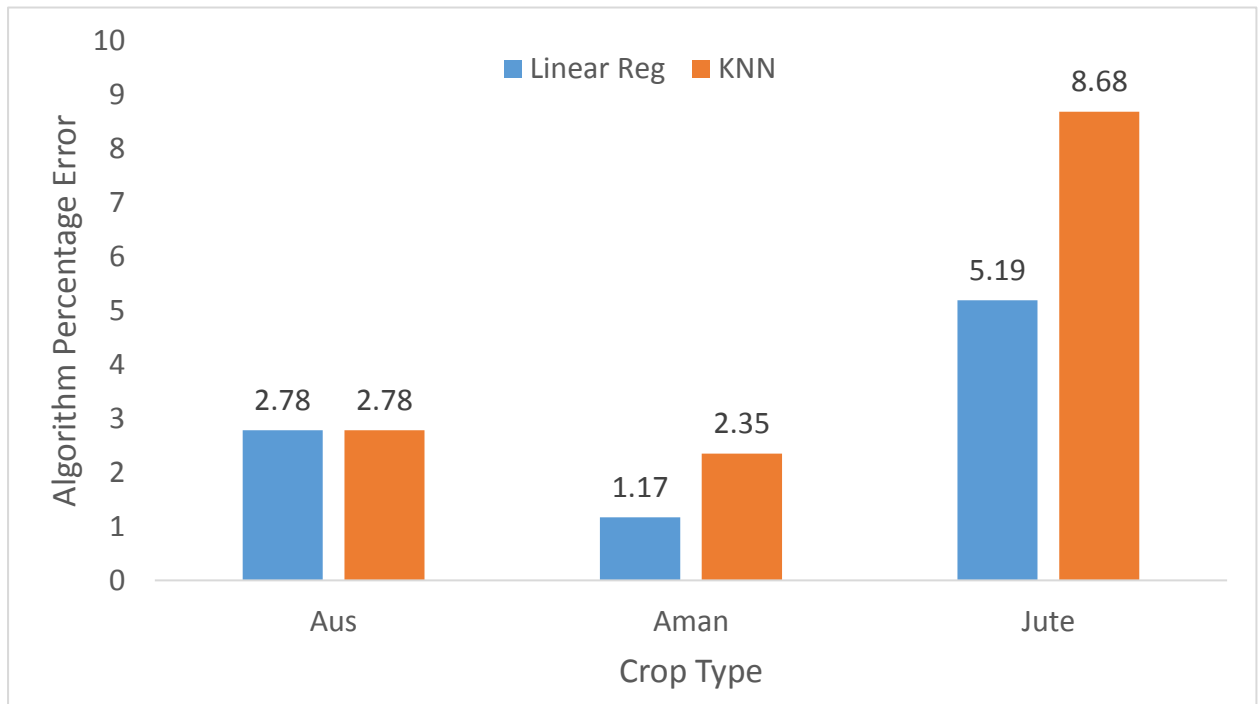


Figure 4.2a: Resulting Percentage Error of K-NN and Linear Regression for Bogra in 2012

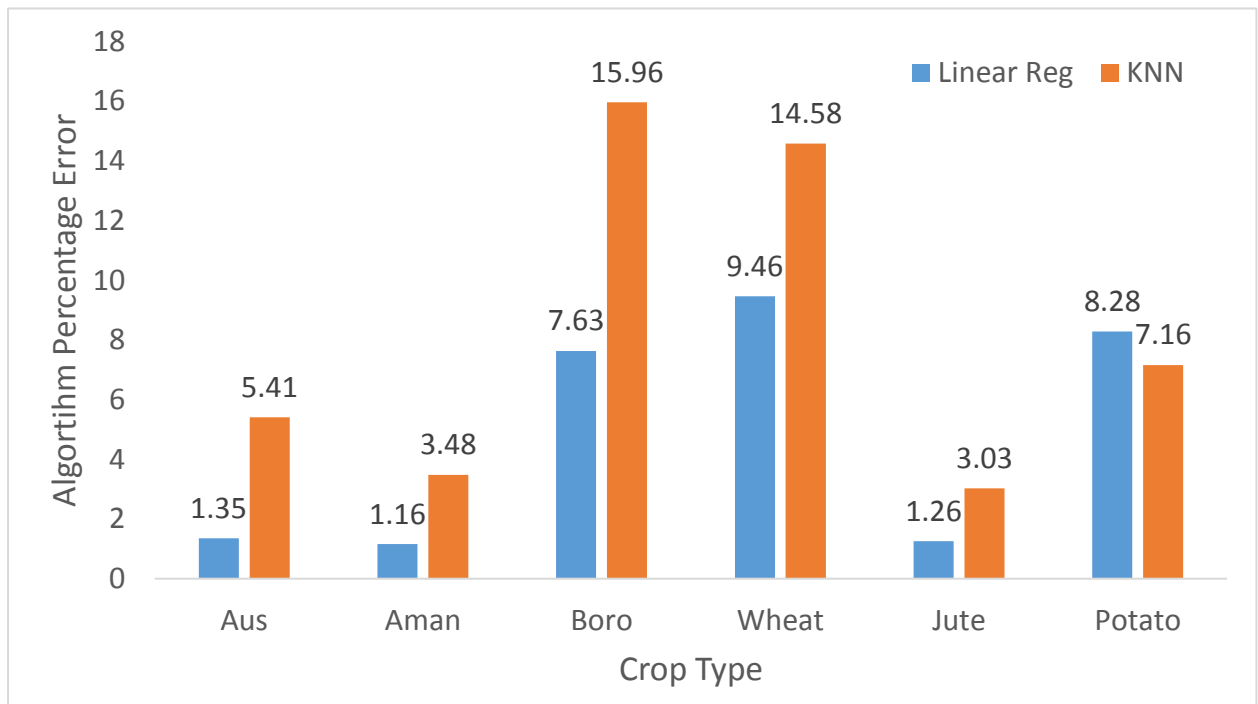


Figure 4.2b: Resulting Percentage Error of K-NN and Linear Regression for Bogra in 2013



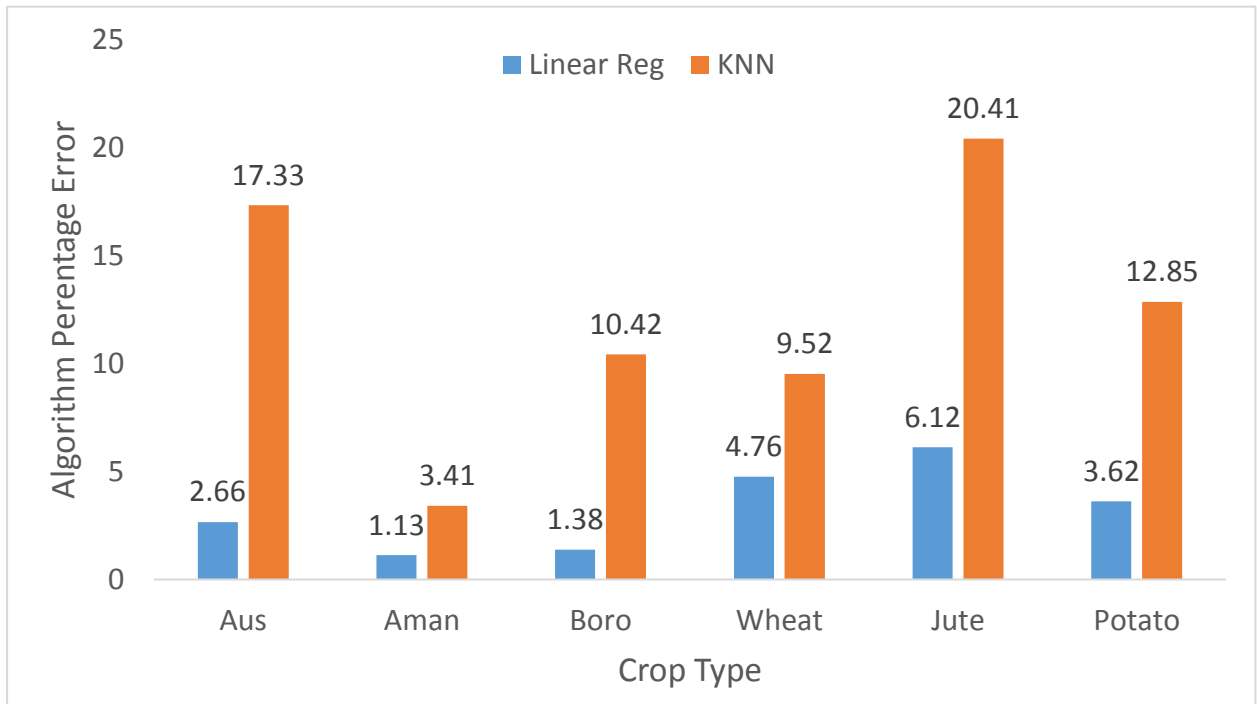


Figure 4.2c: Resulting Percentage Error of K-NN and Linear Regression for Bogra in 2014

The result analysis dataset of Bogra for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of Aman and Jute better than KNN algorithm. And KNN algorithm predicts yield rate of Aus as much as Linear Regression algorithm does.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of the crop types - Aus, Aman, Boro, Wheat and Jute better than KNN algorithm. And KNN algorithm predicts yield rate of Potato better than Linear Regression algorithm.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm as well.

Reason behind this type of result in the year 2013, is that the cropping time-period of Potato (month of November to month of February) in 2013 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned year for the mentioned crop – Potato.

### 4.3 Result Analysis for Comilla Region

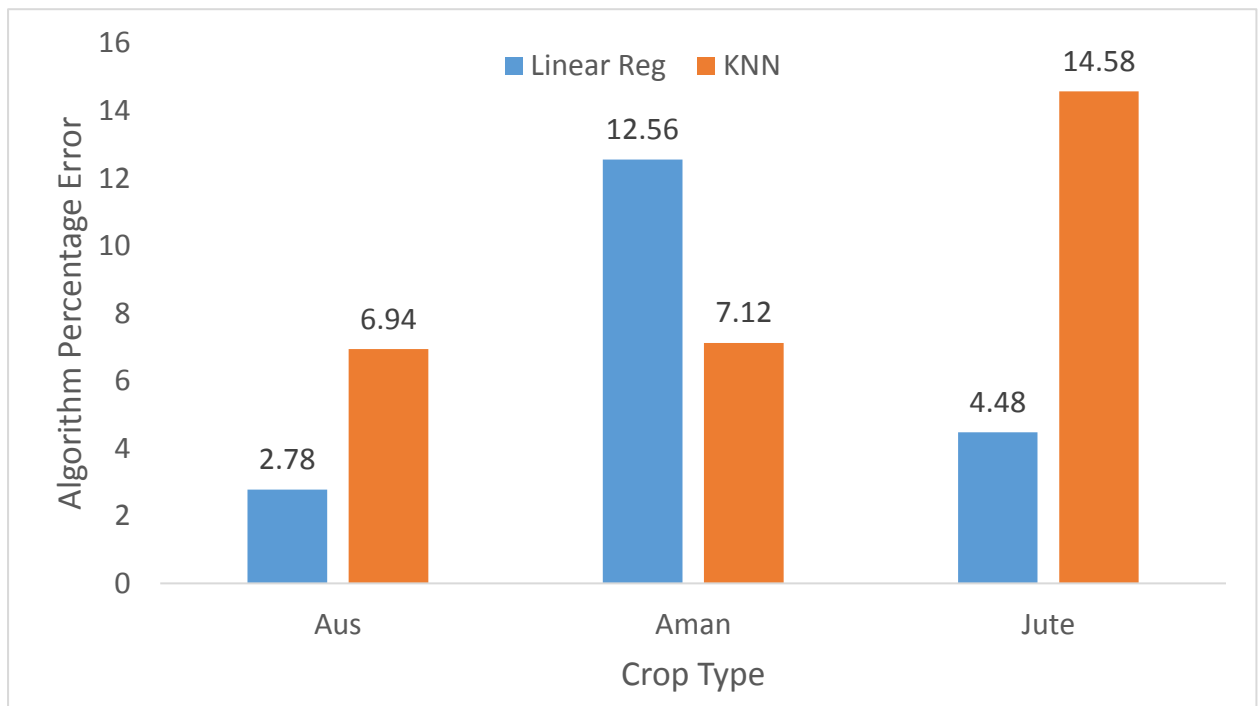


Figure 4.3a: Resulting Percentage Error of K-NN and Linear Regression for Comilla in 2012

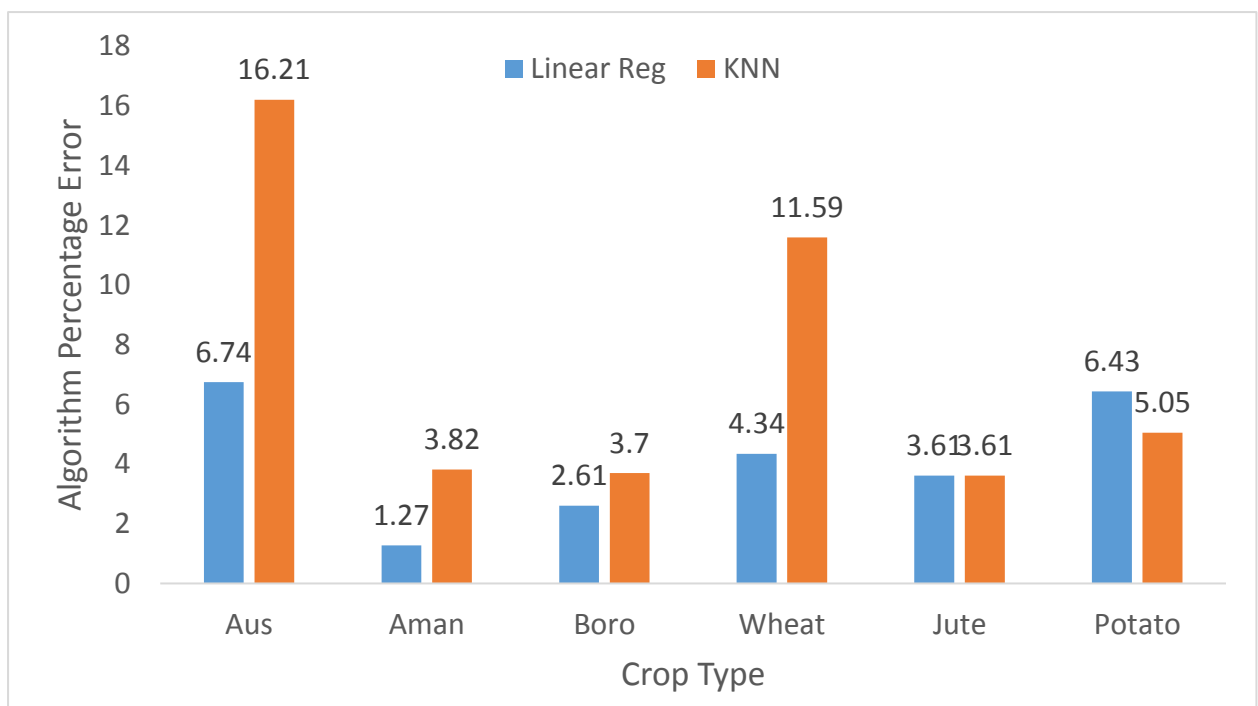


Figure 4.3b: Resulting Percentage Error of K-NN and Linear Regression for Comilla in 2013

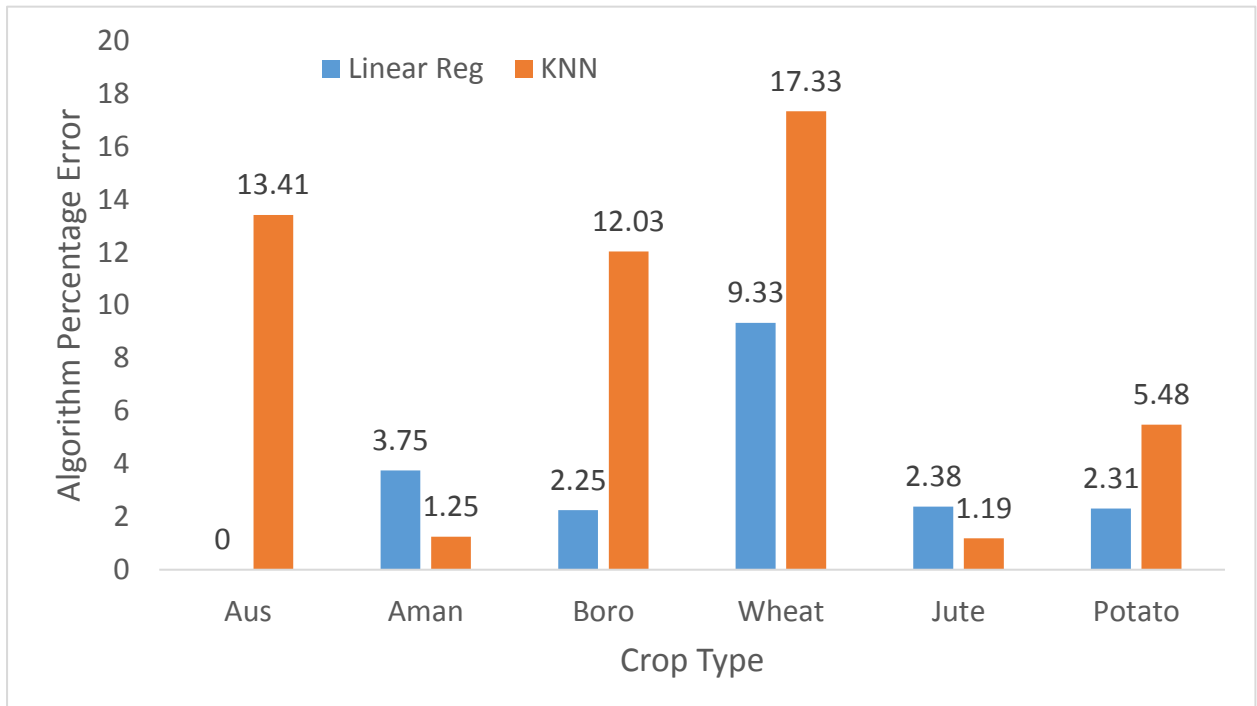


Figure 4.3c: Resulting Percentage Error of K-NN and Linear Regression for Comilla in 2014

The result analysis dataset of Comilla for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of Aus and Jute better than KNN algorithm. And KNN algorithm predicts yield rate of Aman better than Linear Regression algorithm does.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of the crop types - Aus, Aman, Boro and Wheat better than KNN algorithm. And KNN algorithm predicts yield rate of Potato better than Linear Regression algorithm, and for Jute, both algorithm's performance was same.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of the crop types - Aus, Boro, Wheat and Potato better than KNN algorithm. And KNN algorithm predicts yield rate of the crop types - Aman and Jute better than Linear Regression algorithm.

Reason behind this type of result in the years 2012, 2013 and 2014 is that the cropping time-period of Aman (month of June to month of December) in 2012, Potato (month of November to month of February) in 2013 and Aman (month of June to month of December) and Jute (month of April to month of July) in 2014 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11).

Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned years for the mentioned crops.

#### 4.4 Result Analysis for Dhaka region

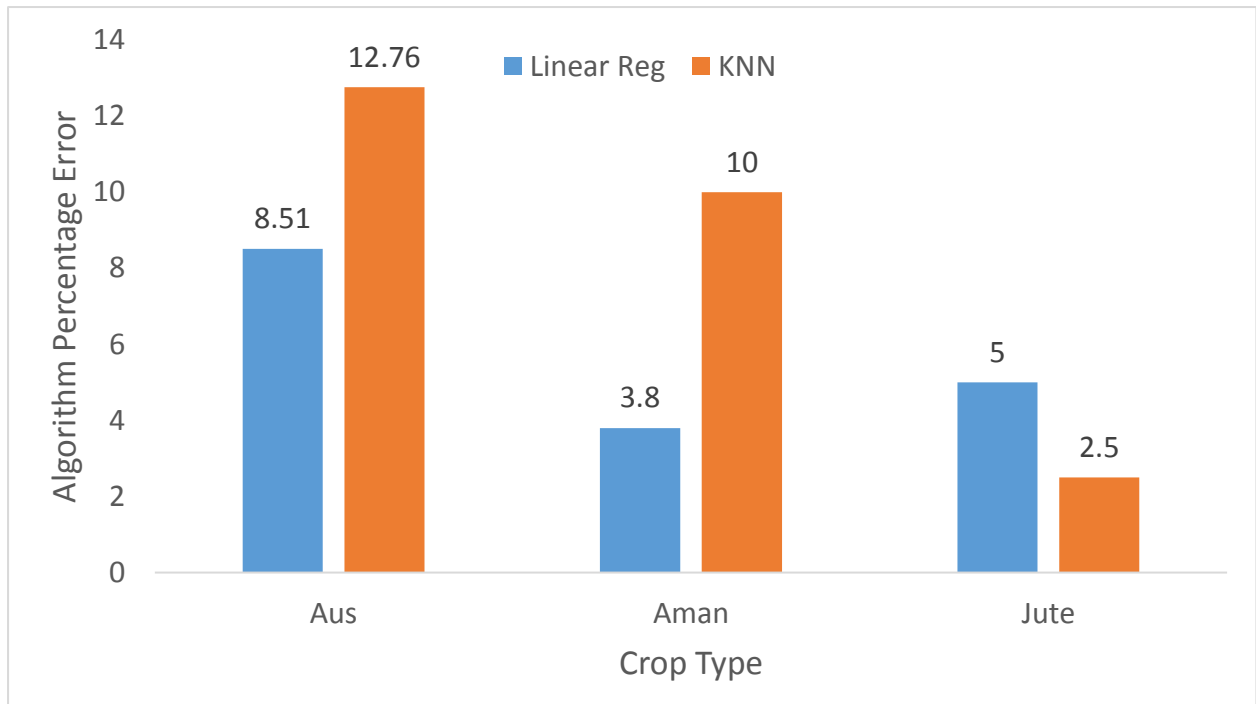


Figure 4.4a: Resulting Percentage Error of K-NN and Linear Regression for Dhaka in 2012

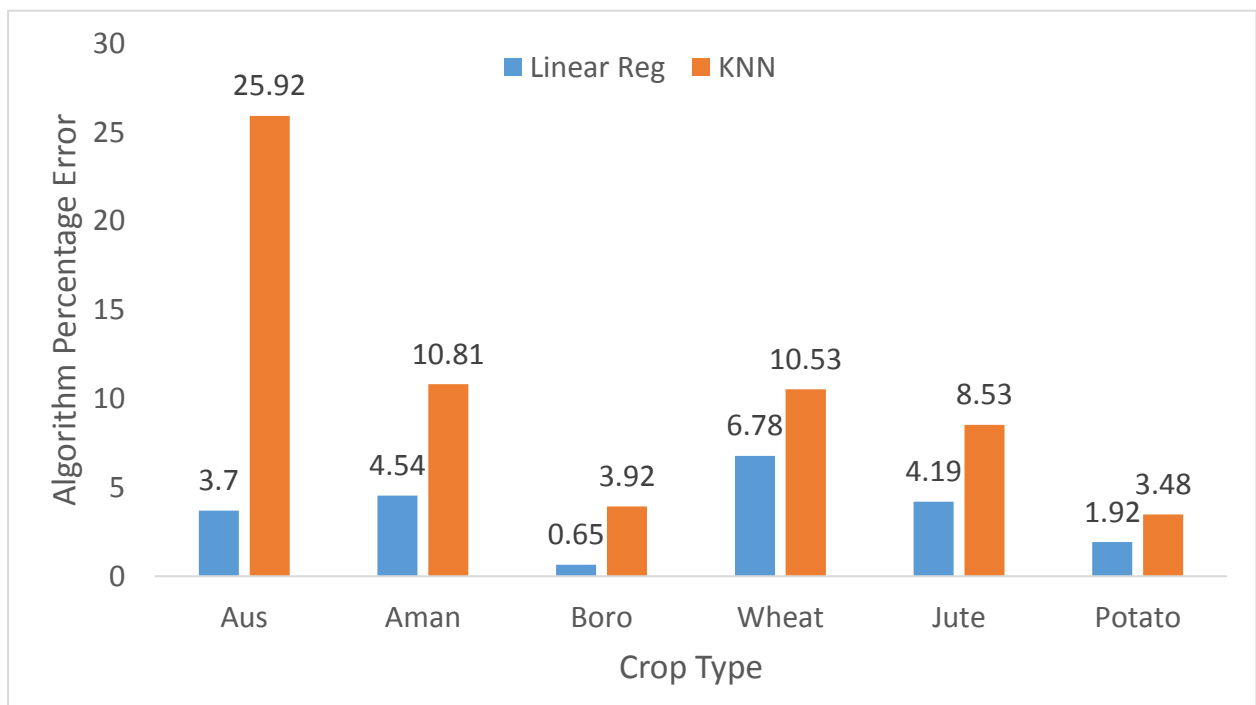


Figure 4.4b: Resulting Percentage Error of K-NN and Linear Regression for Dhaka in 2013

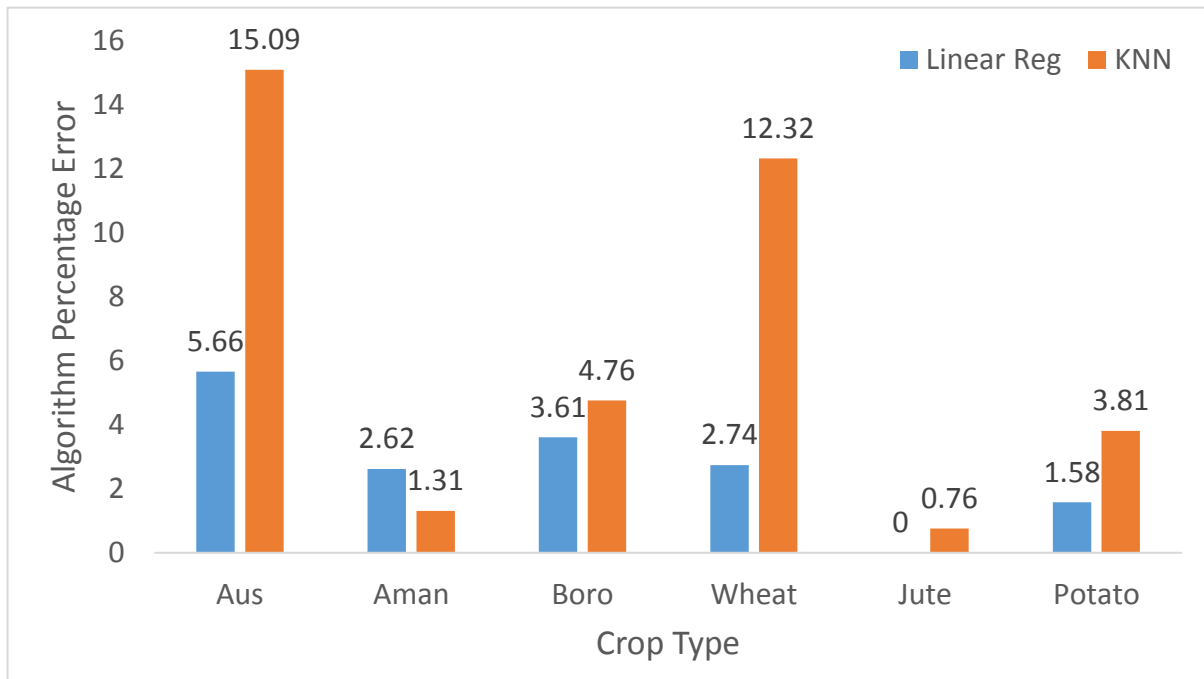


Figure 4.4c: Resulting Percentage Error of K-NN and Linear Regression for Dhaka in 2014

The result analysis dataset of Dhaka for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of Aus and Aman better than KNN algorithm. And KNN algorithm predicts yield rate of Jute better than Linear Regression algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of Aus, Boro, Wheat, Jute and Potato better than KNN algorithm. However, KNN algorithm predicts yield rate of Aman better than Linear Regression algorithm.

Reason behind this type of result in the years 2012 and 2014, is that the cropping time-period of Jute (month of April to month of July) in 2012 and Aman (month of June to month of December) in 2014 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned years for the mentioned crop.

#### 4.5 Result Analysis for Dinajpur region

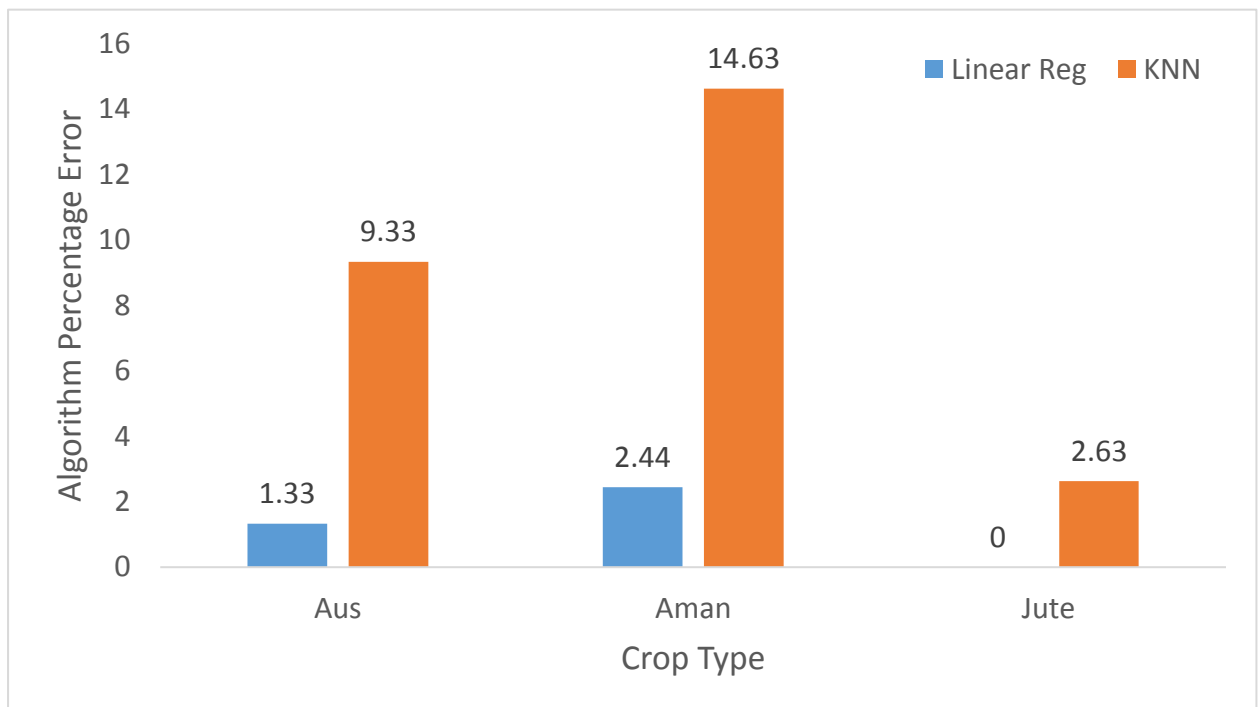


Figure 4.5a: Resulting Percentage Error of K-NN and Linear Regression for Dinajpur in 2012

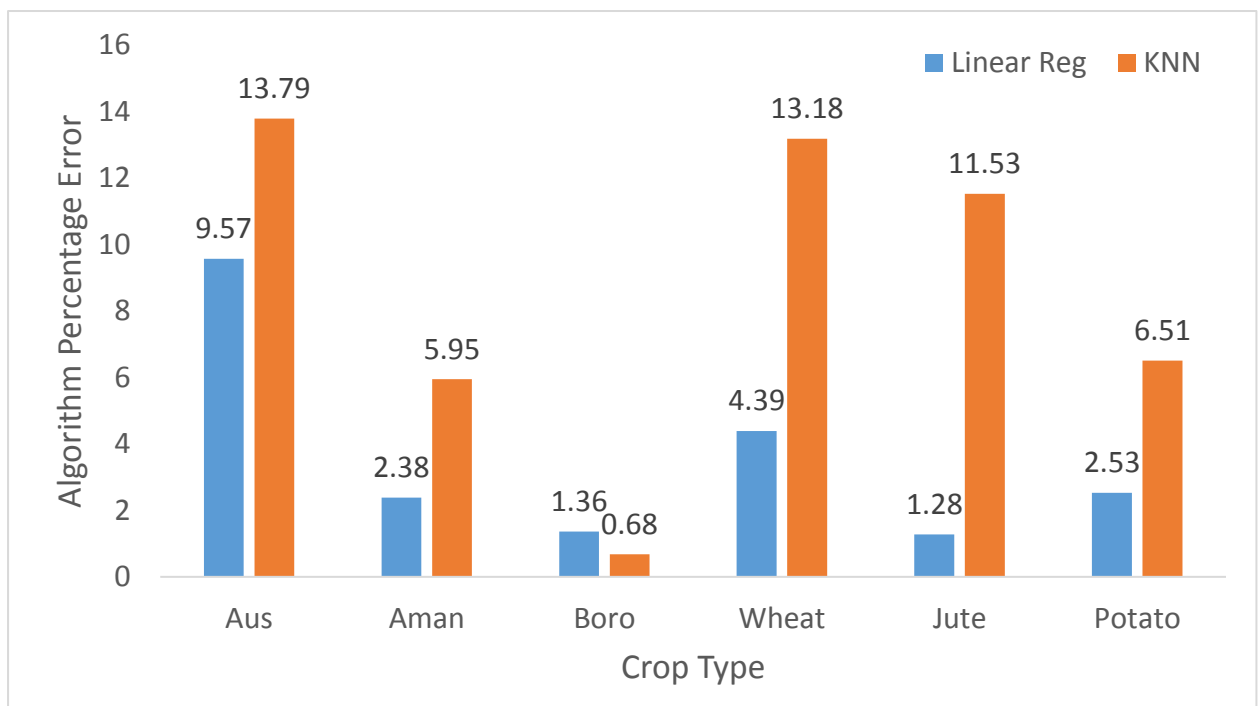


Figure 4.5b: Resulting Percentage Error of K-NN and Linear Regression for Dinajpur in 2013

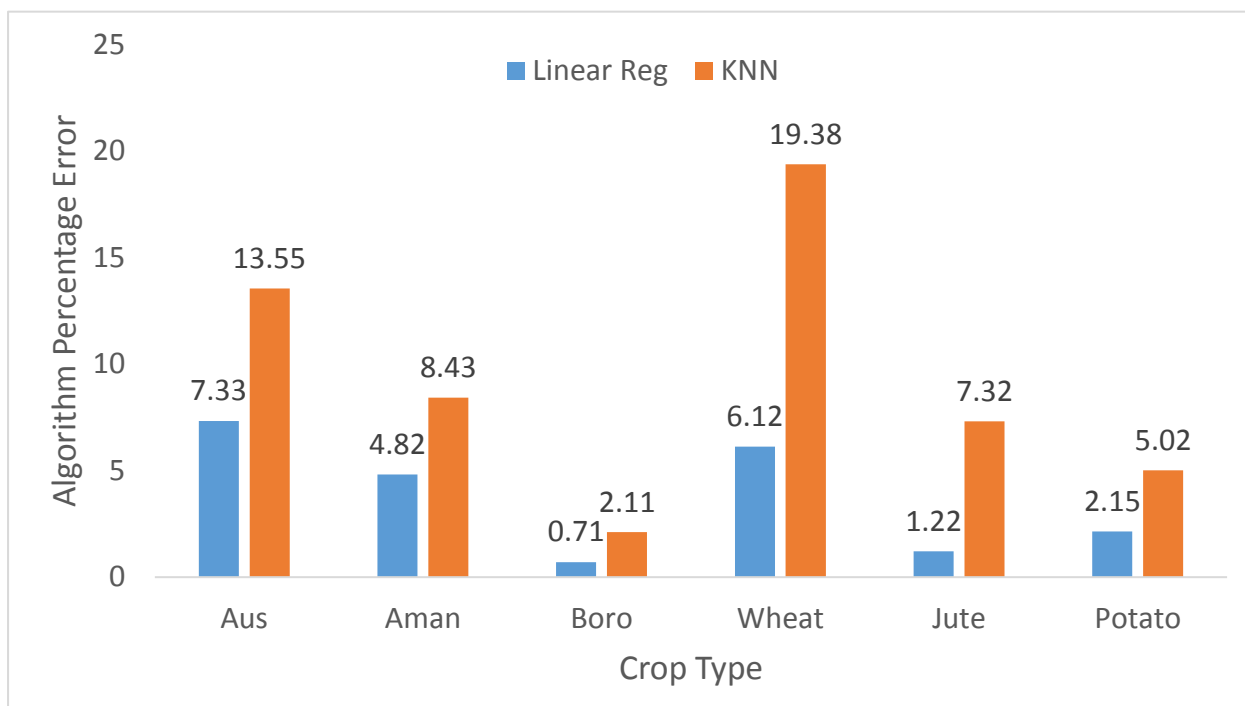


Figure 4.5c: Resulting Percentage Error of K-NN and Linear Regression for Dinajpur in 2014

The result analysis dataset of Dinajpur for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of all the crop types – Aus, Aman and Jute better than KNN algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of the crop types – Aus, Aman, Wheat, Jute and Potato better than KNN algorithm. And KNN algorithm predicts yield rate of Boro better than Linear Regression algorithm.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm.

Reason behind this type of result in the year 2013, is that the cropping time-period of Boro (month of October to month of May) in 2013 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned year for the mentioned crop.

#### 4.6 Result Analysis for Faridpur region

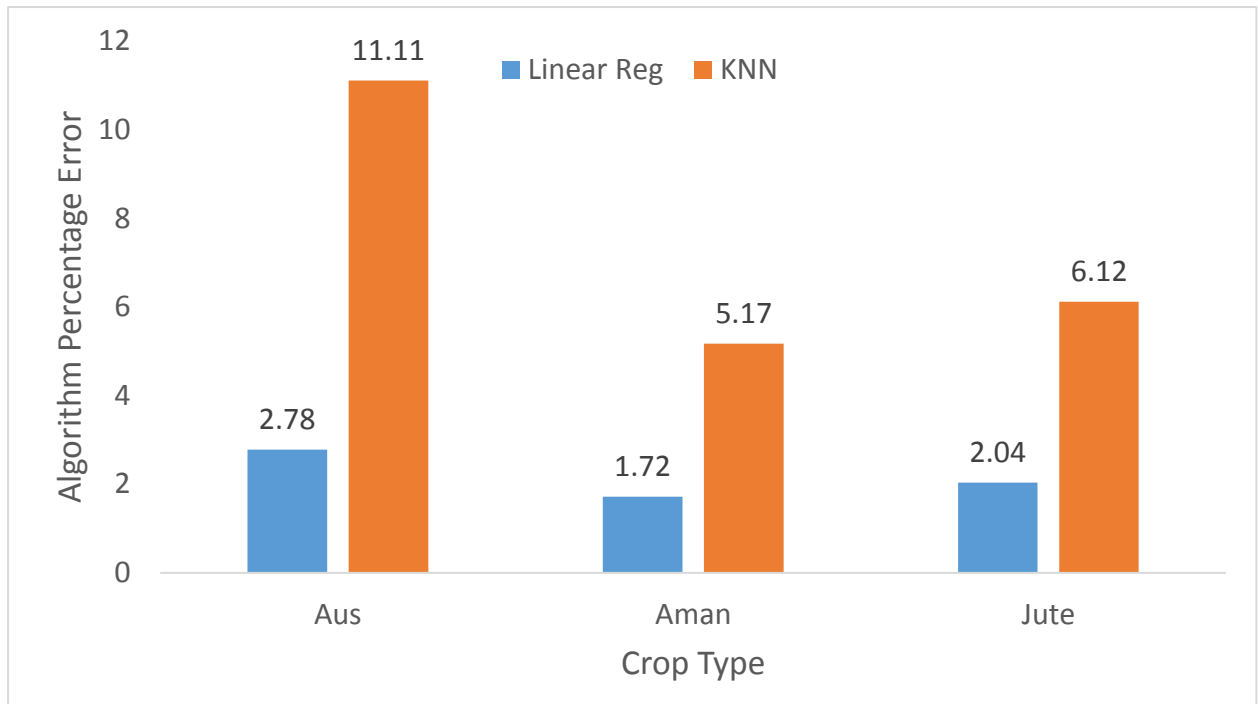


Figure 4.6a: Resulting Percentage Error of K-NN and Linear Regression for Faridpur in 2012

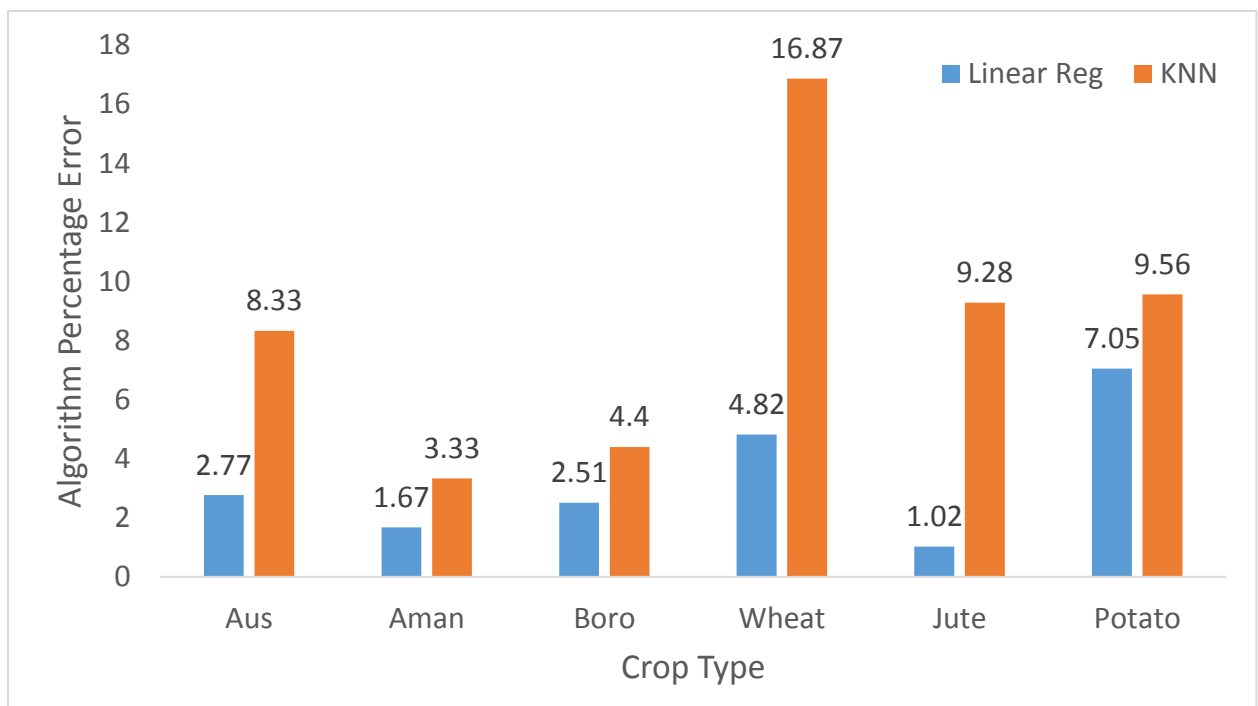


Figure 4.6b: Resulting Percentage Error of K-NN and Linear Regression for Faridpur in 2013



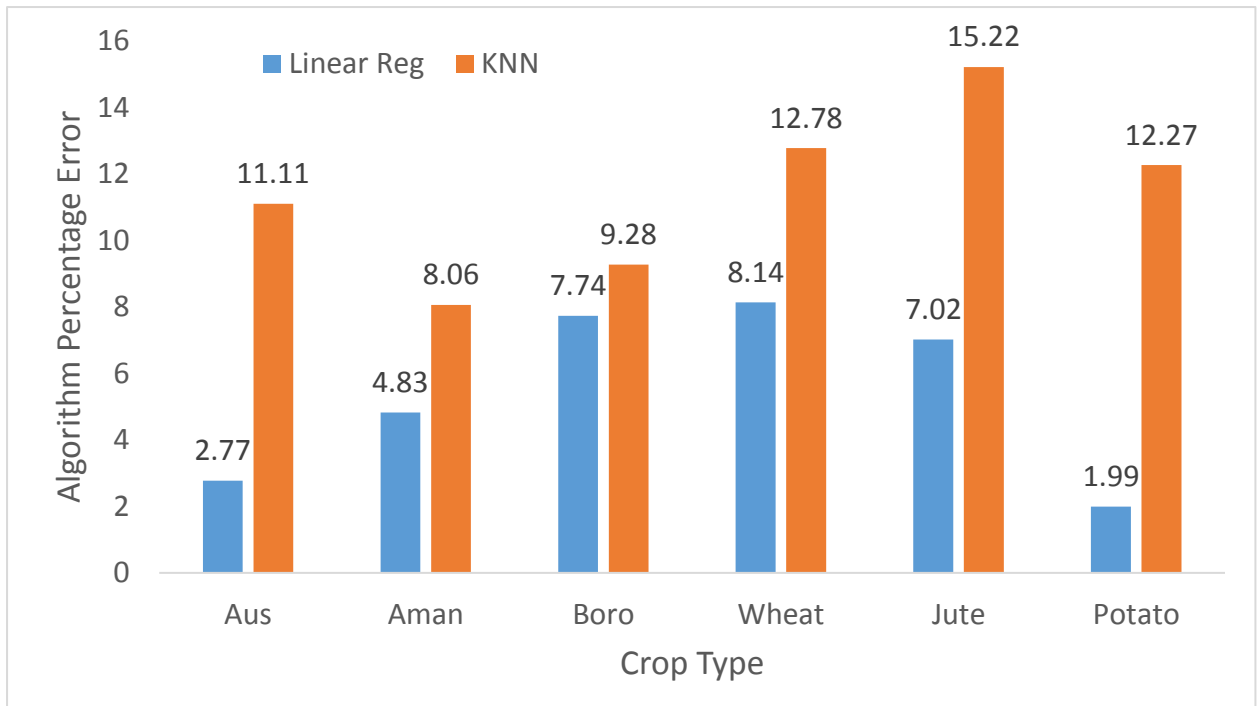


Figure 4.6c: Resulting Percentage Error of K-NN and Linear Regression for Faridpur in 2014

The result analysis dataset of Faridpur for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of all the crop types – Aus, Aman and Jute better than KNN algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm as well.

Again, in the year 2014, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm as well.

Reason behind this type of result in the testing years, is that the cropping time-period of all the crop types throughout all the 3 years had rainfall and temperature amount very consistent and near to the Linear Regression line of the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm predicted more accurately than KNN algorithm in all the years for all the crops.

#### 4.7 Result Analysis for Khulna region

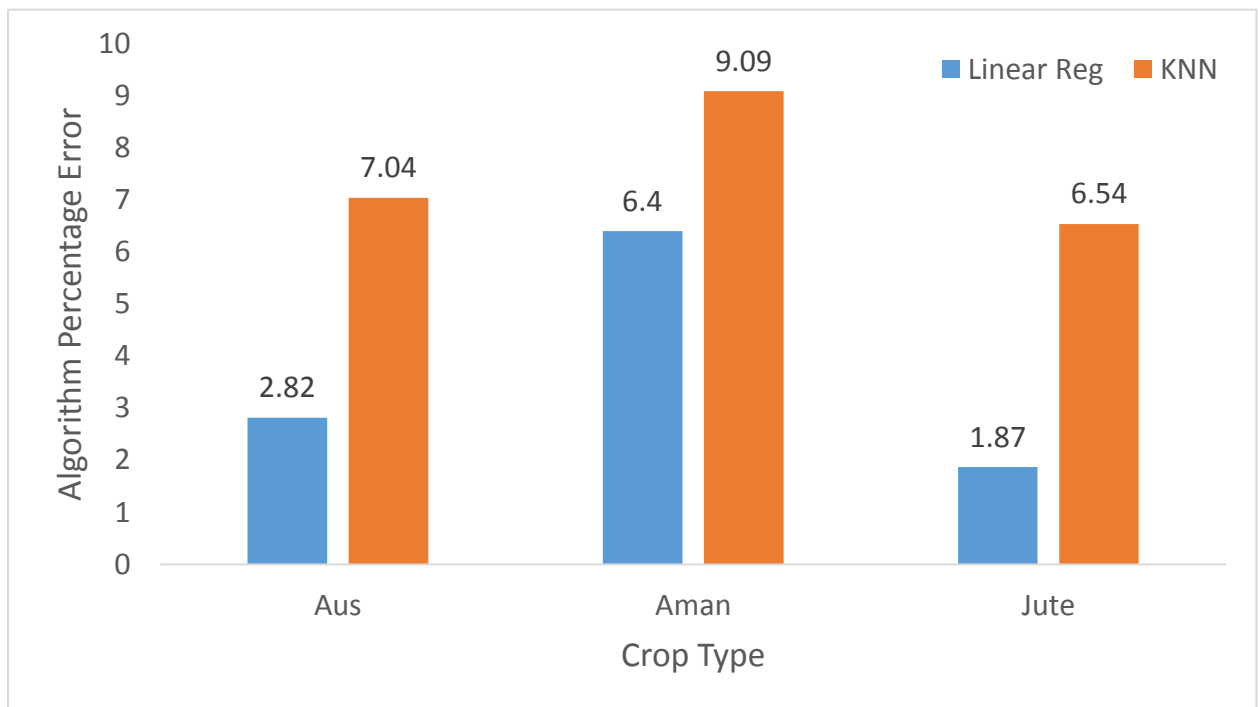


Figure 4.7a: Resulting Percentage Error of K-NN and Linear Regression for Khulna in 2012

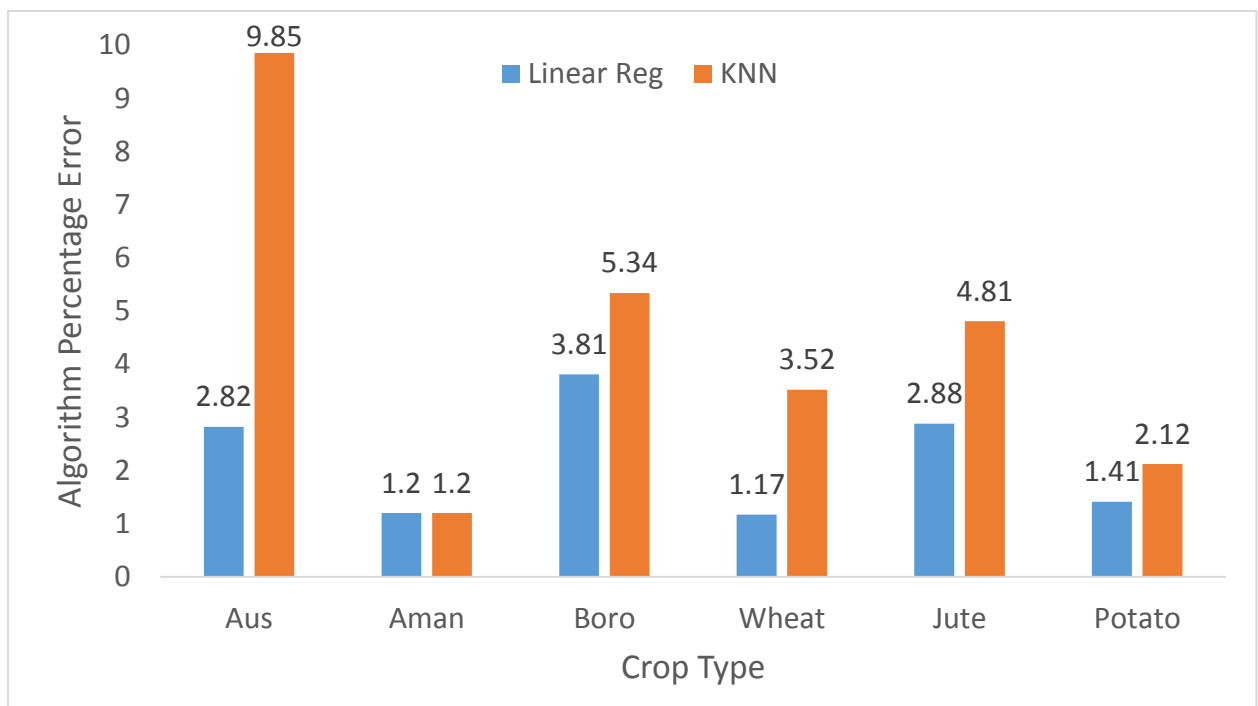


Figure 4.7b: Resulting Percentage Error of K-NN and Linear Regression for Khulna in 2013

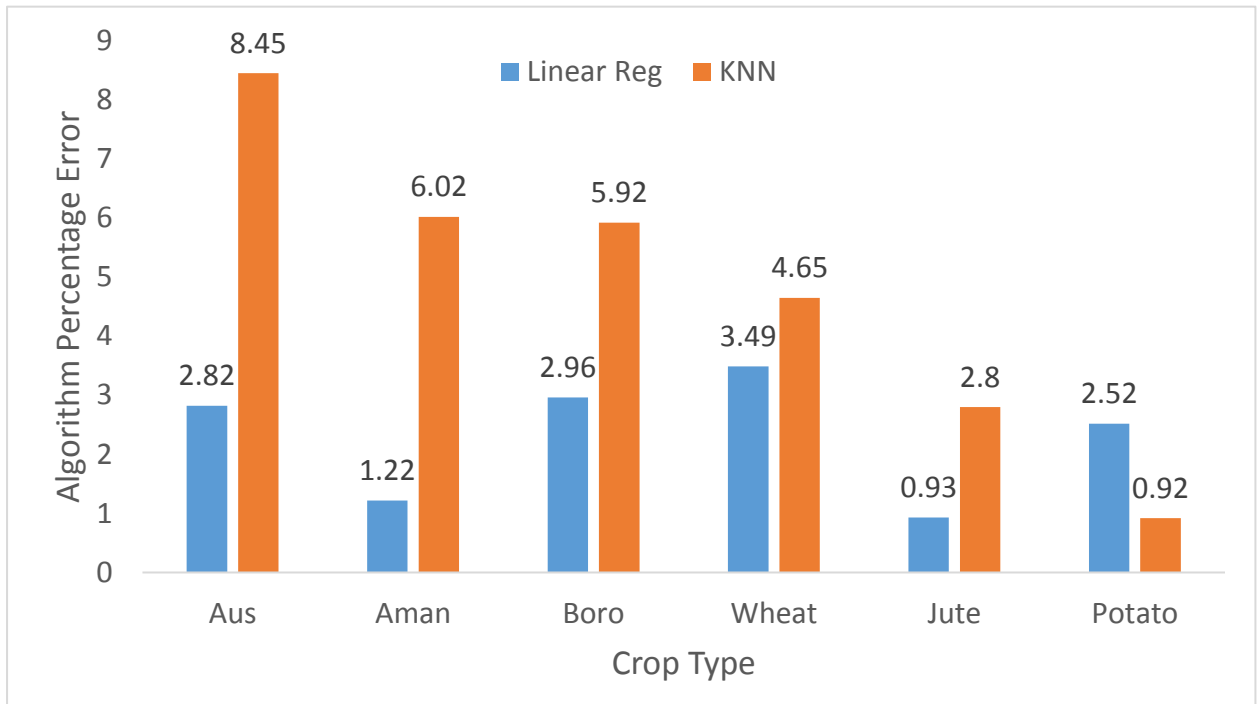


Figure 4.7c: Resulting Percentage Error of K-NN and Linear Regression for Khulna in 2014

The result analysis dataset of Khulna for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of all the crop types – Aus, Aman and Boro better than KNN algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of the crop types – Aus, Boro, Wheat, Jute and Potato better than KNN algorithm. However, KNN algorithm predicts yield rate of Aman as much as Linear Regression algorithm does.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of Aus, Boro, Wheat, Jute and Potato better than KNN algorithm. However, KNN algorithm predicts yield rate of Potato better than Linear Regression algorithm.

Reason behind this type of result in the years 2013 and 2014, is that the cropping time-period of Aman (month of June to month of December) in 2013 and Potato (month of November to month of February) in 2014 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned years for the mentioned crop.

#### 4.8 Result Analysis for Rajshahi region

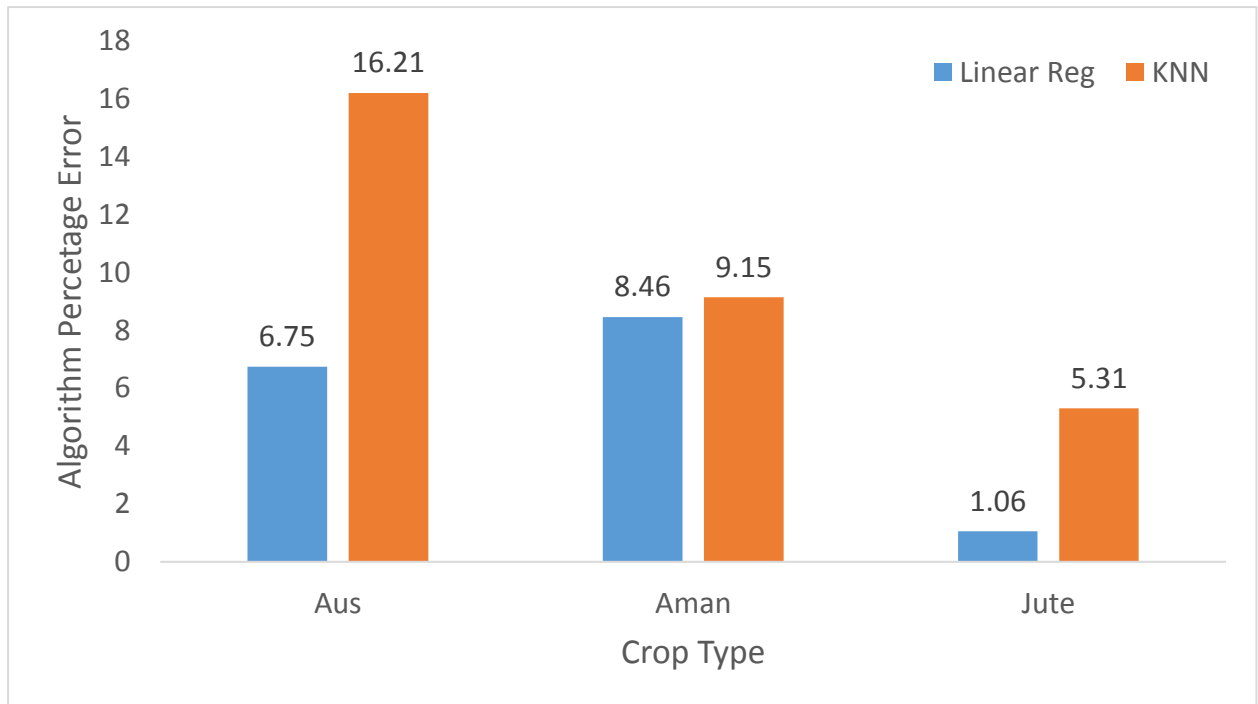


Figure 4.8a: Resulting Percentage Error of K-NN and Linear Regression for Rajshahi in 2012

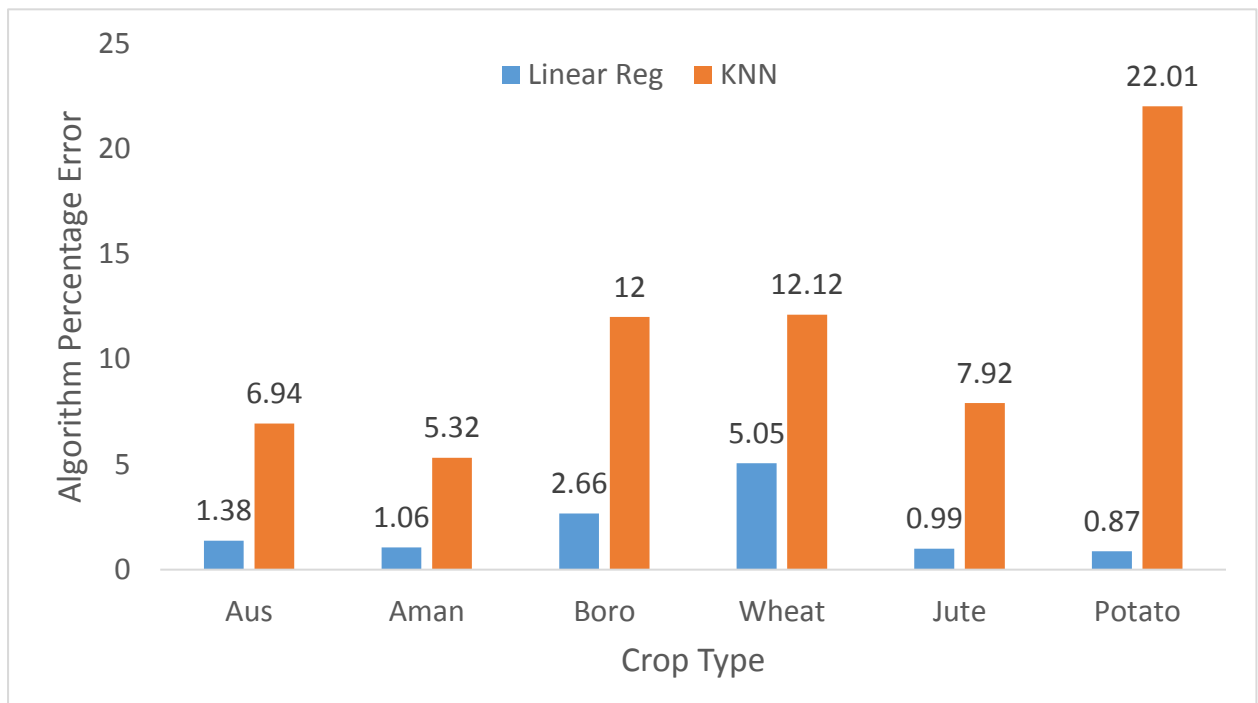


Figure 4.8b: Resulting Percentage Error of K-NN and Linear Regression for Rajshahi in 2013

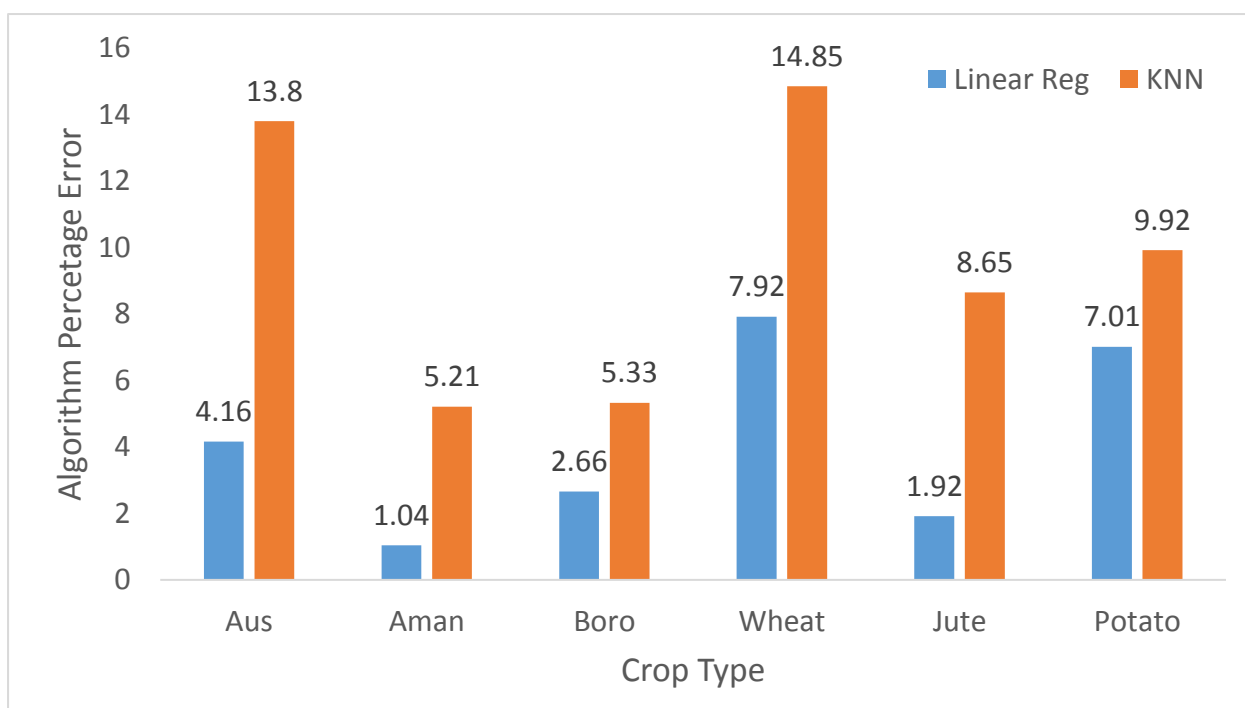


Figure 4.8c: Resulting Percentage Error of K-NN and Linear Regression for Rajshahi in 2014

The result analysis dataset of Rajshahi for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of all the crop types – Aus, Aman and Jute better than KNN algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm.

Again, in the year 2014, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm.

Reason behind this type of result in the testing years, is that the cropping time-period of all the crop types throughout all the 3 years had rainfall and temperature amount very consistent and near to the Linear Regression line of the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm predicted more accurately than KNN algorithm in all the years for all the crops.

#### 4.9 Result Analysis for Rangpur region

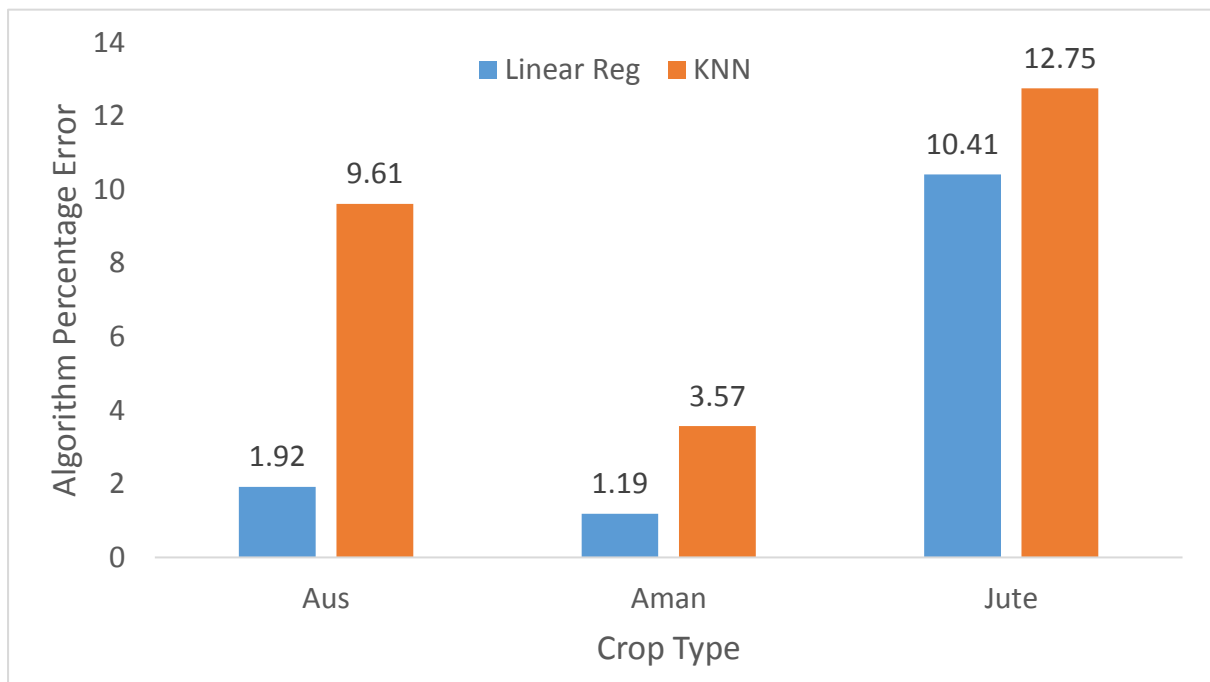


Figure 4.9a: Resulting Percentage Error of K-NN and Linear Regression for Rangpur in 2012

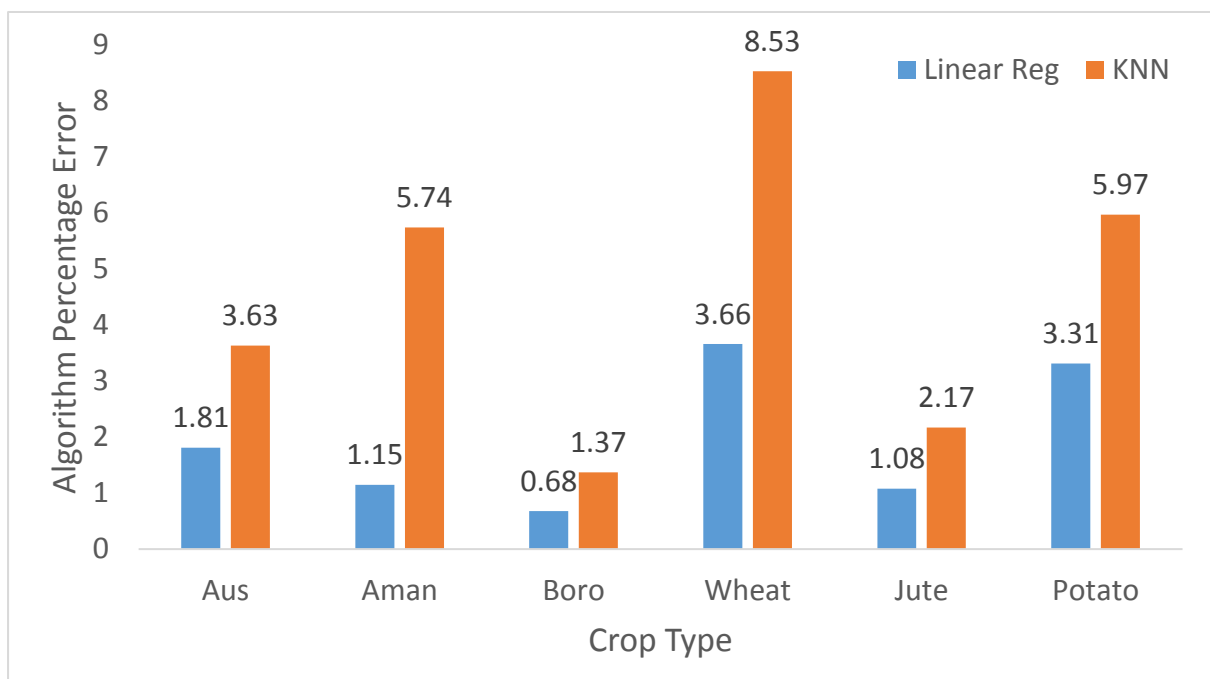


Figure 4.9b: Resulting Percentage Error of K-NN and Linear Regression for Rangpur in 2013

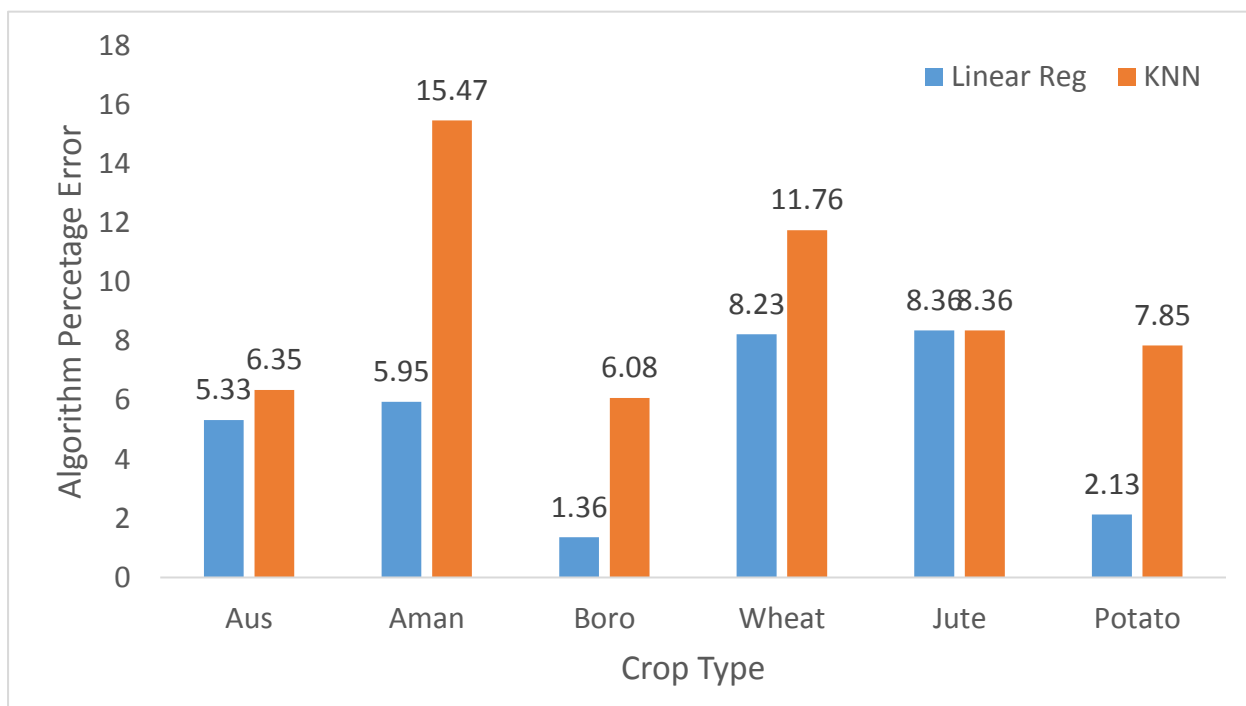


Figure 4.9c: Resulting Percentage Error of K-NN and Linear Regression for Rangpur in 2014

The result analysis dataset of Dhaka for 3 years (2012-14) is given above.

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of all the crop types – Aus, Aman and Jute better than KNN algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro, Wheat, Jute and Potato better than KNN algorithm.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of Aus, Aman, Boro, Wheat and Potato better than KNN algorithm. However, KNN algorithm predicts yield rate of Jute as much as Linear Regression algorithm.

Reason behind this type of result in the year 2014, is that the cropping time-period of Jute (month of April to month of July) in 2014 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned year for the mentioned crop.

#### 4.10 Result Analysis for Sylhet region

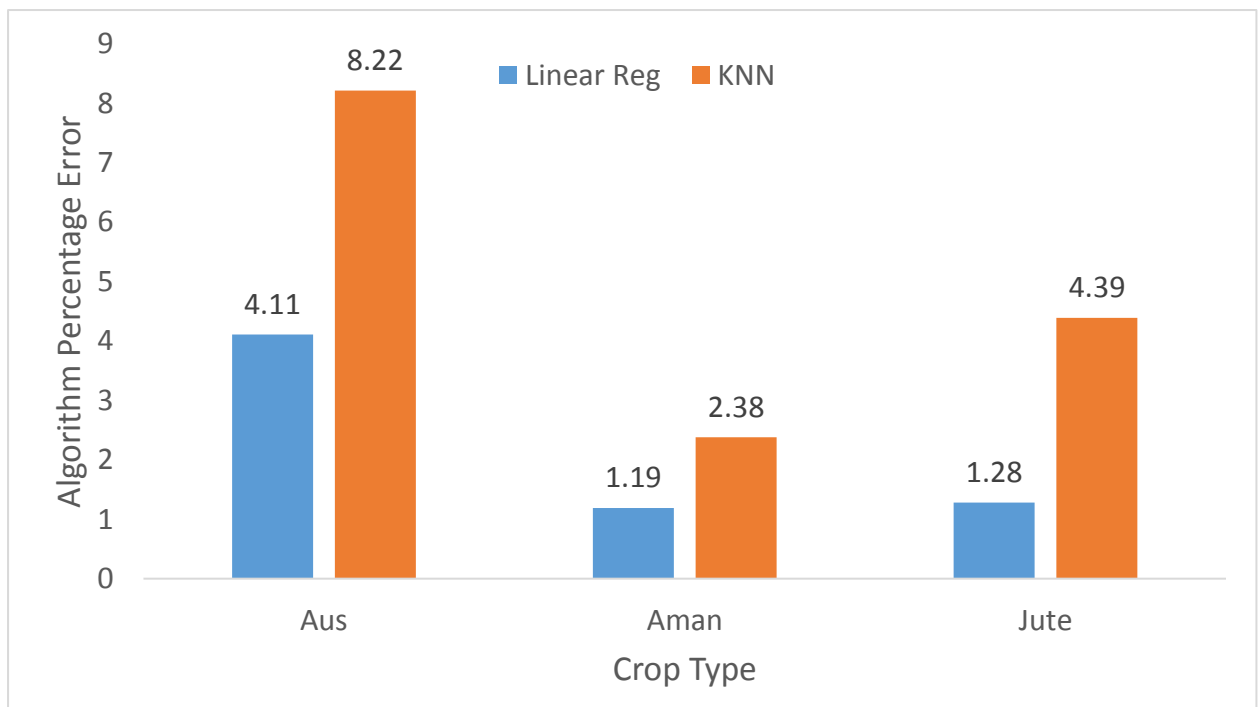


Figure 4.10a: Resulting Percentage Error of K-NN and Linear Regression for Sylhet in 2012

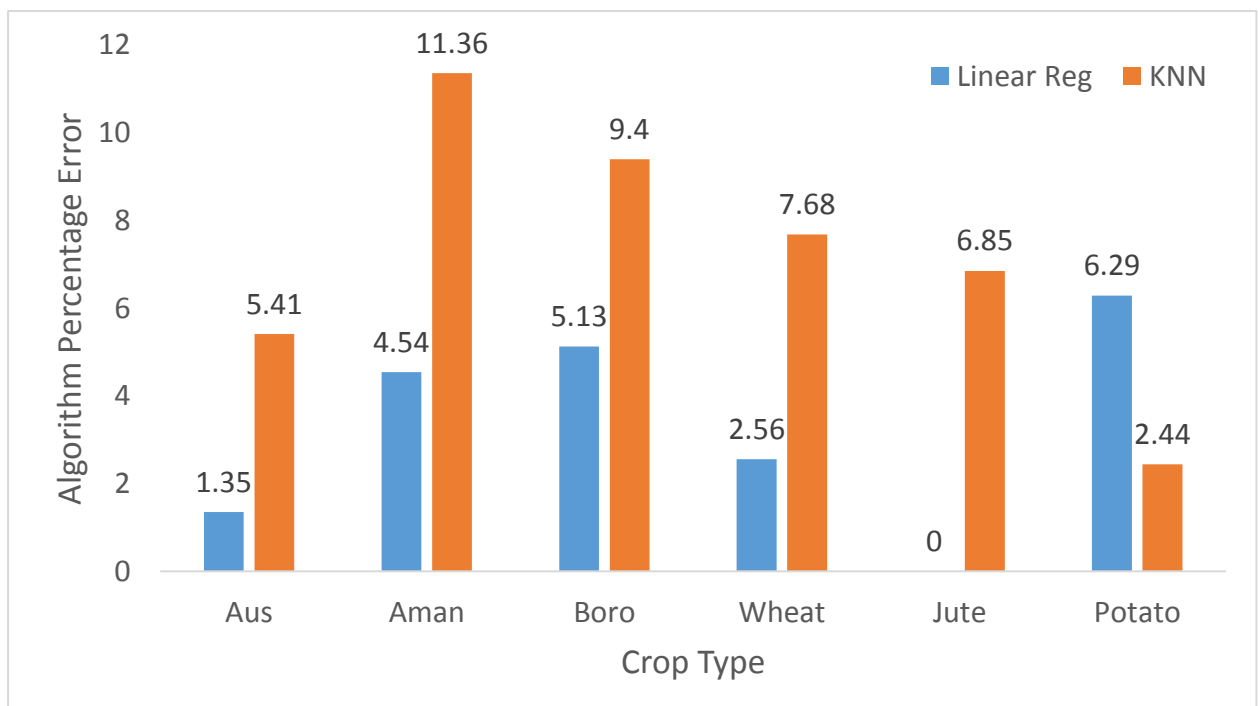


Figure 4.10b: Resulting Percentage Error of K-NN and Linear Regression for Sylhet in 2013



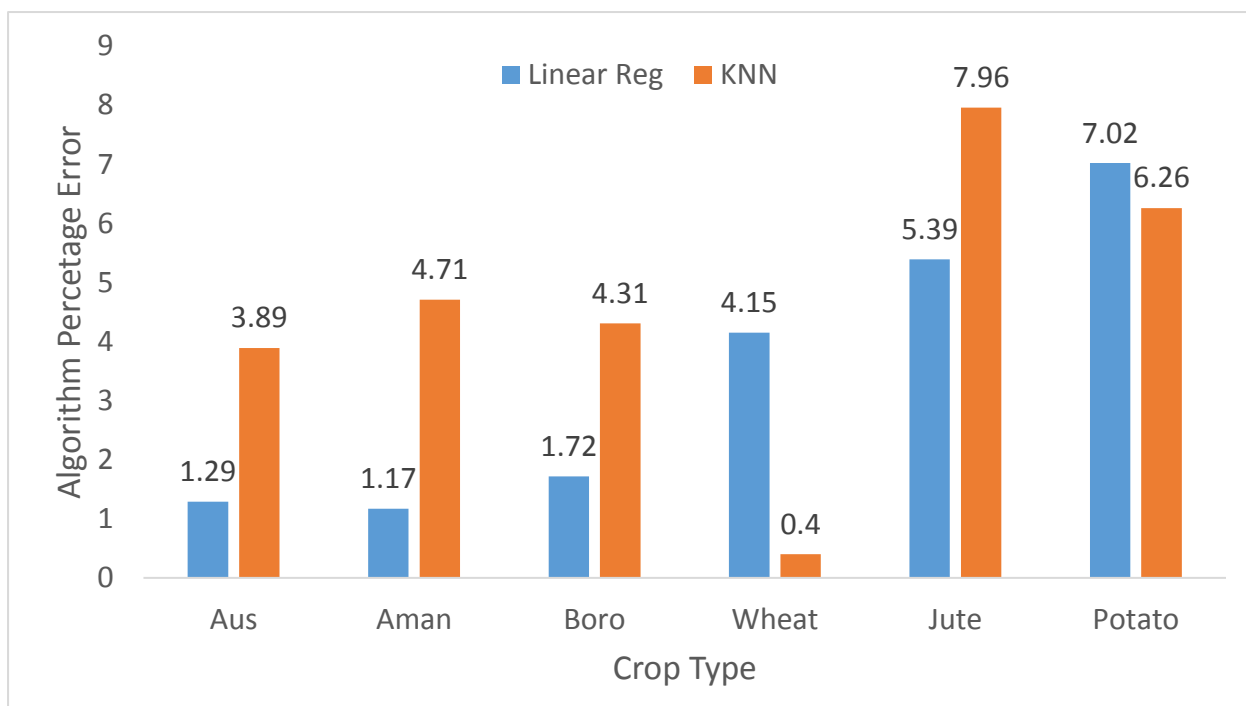


Figure 4.10c: Resulting Percentage Error of K-NN and Linear Regression for Sylhet in 2014

As we can see, in the year 2012, Linear Regression algorithm predicts the probable yield rate of all the crop types – Aus, Aman and Jute better than KNN algorithm.

In the year 2013, Linear Regression algorithm predicts the probable yield rate of the crop types – Aus, Aman, Boro, Wheat and Jute better than KNN algorithm. And KNN algorithm predicts yield rate of Potato better than Linear Regression algorithm.

In the year 2014, Linear Regression algorithm predicts the probable yield rate of all the crop types - Aus, Aman, Boro and Jute better than KNN algorithm. However, KNN algorithm predicts yield rate of Wheat and Potato better than Linear Regression algorithm.

Reason behind this type of result in the years 2013 and 2014, is that the cropping time-period of Potato (month of November to month of February) in 2013 and cropping time-period of Wheat (month of November to month of April) and Potato in 2014 had rainfall and temperature amount varied a little more than the usual average temperature and rainfall amount in the learning years (2000-11). Because Linear Regression algorithm gives better result when testing variable amounts are closer to the learning variables' average amount, this algorithm could not predict more accurately than KNN algorithm in the mentioned year for the mentioned crop.

#### 4.11 Data Analysis before and after adding synthetic data, removing outliers

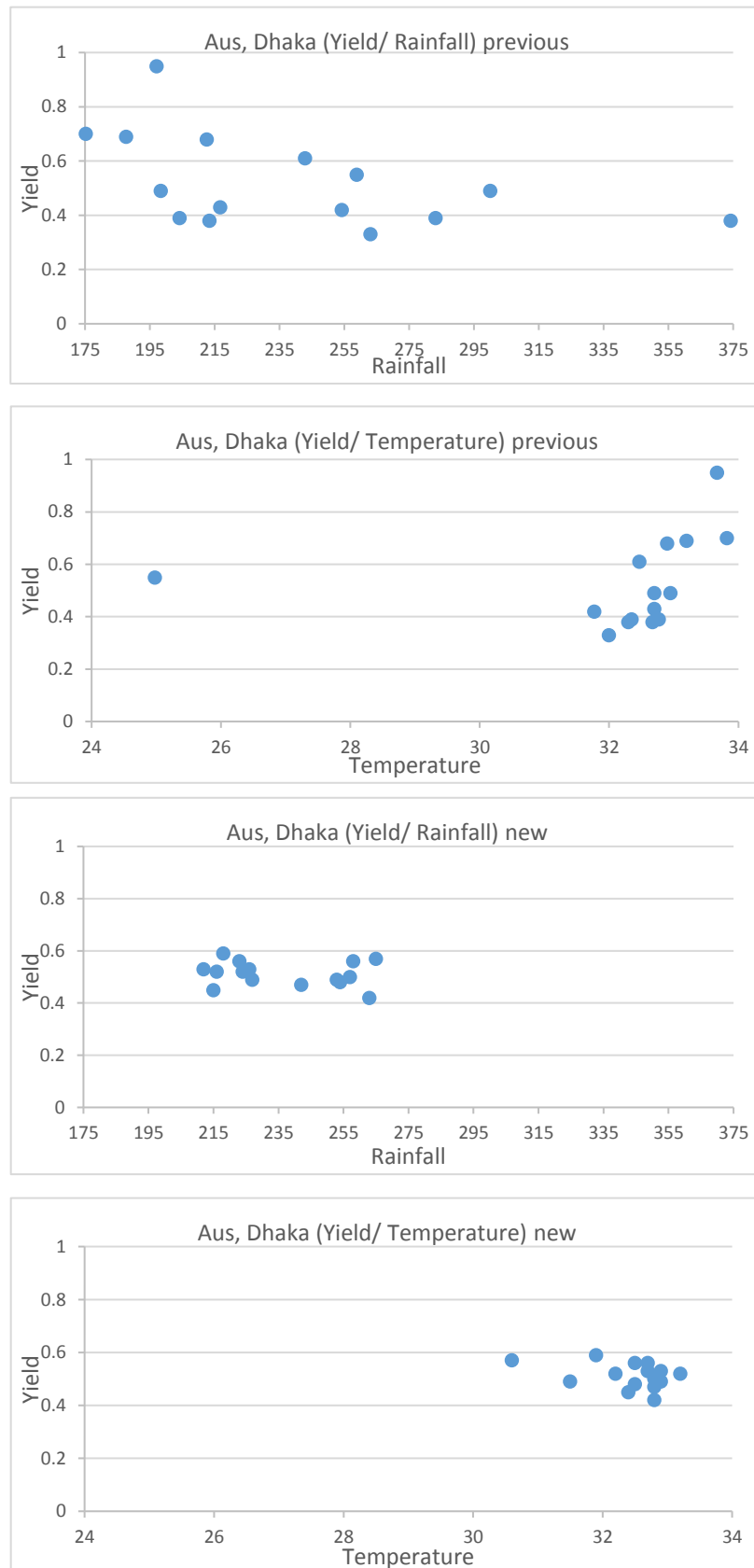


Figure 4.11a-d: Data Analysis for Aus of Dhaka region

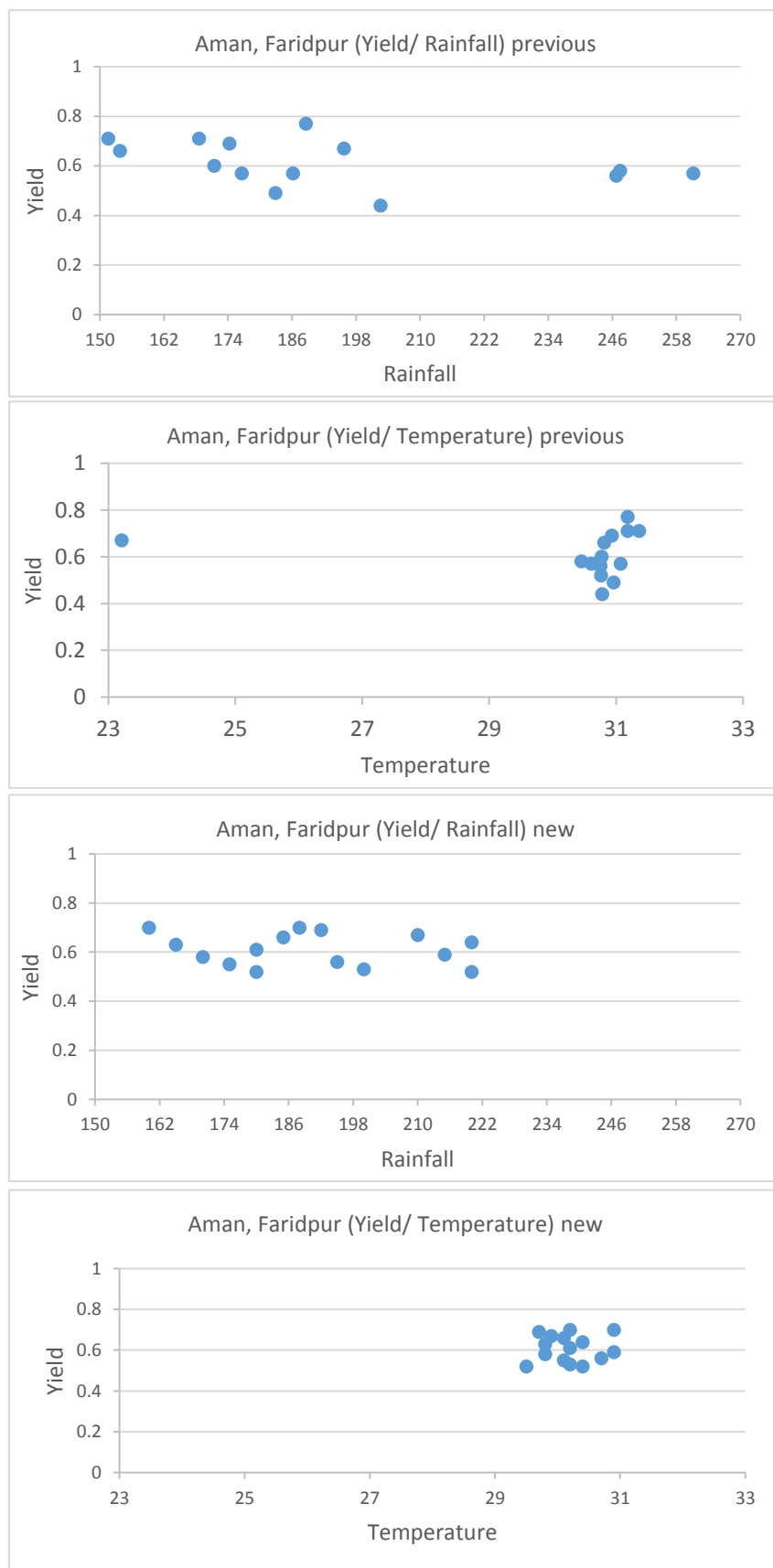


Figure 4.11e-h: Data Analysis for Aman of Faridpur region

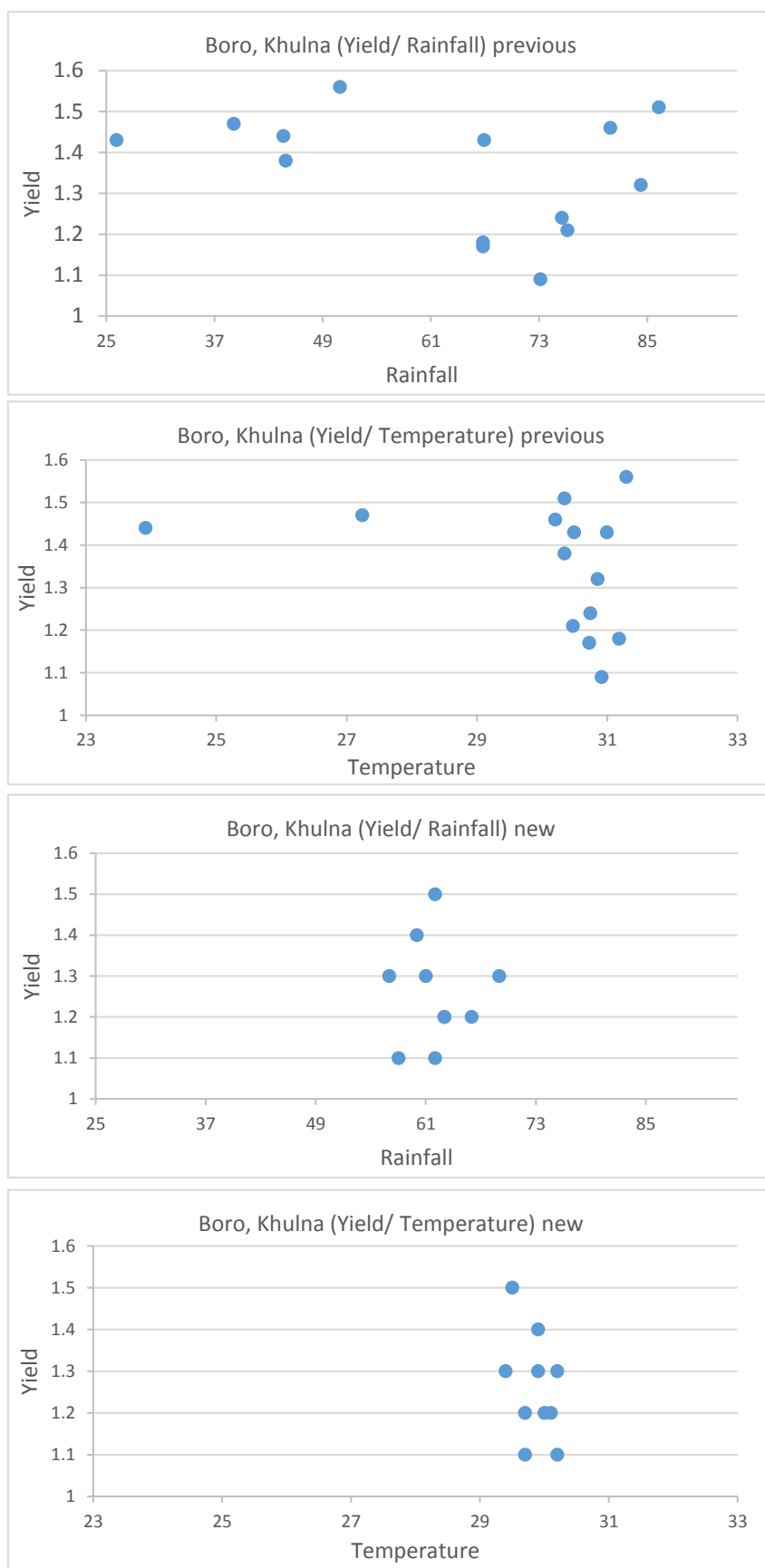


Figure 4.11i-l: Data Analysis for Boro of Khulna region

As we can see, the 4 diagrams (Figure 4.11a to 4.11d) symbolize 2 different graphs for ‘yield vs rainfall’ and ‘yield vs temperature’ measurements for the crop type Aus of region Dhaka. The 2 graphs are made for 2 scenarios – one before adding synthetic data and removing the outliers from calculation and one after that. Same procedure was followed for crop type Aman of region Faridpur (Figure 4.11e to Figure 4.11h) and crop type Boro of region Khulna (Figure 4.11i to Figure 4.11l).

When we see graphs – Figure 4.11a and Figure 4.11b; we can see that of the crop type Aus in the Dhaka region, data points for ‘yield vs rainfall’ and ‘yield vs temperature’ are scattered a little too much. When we consulted our supervisor and the agriculture extension experts about this, they suggested that we could bring some modifications in our algorithms so that it will not consider the outlier/ extremely distant data points and add some synthetic data. By following the suggestion, we found the graphs – Figure 4.11c and Figure 4.11d. In these graphs, we can see that the data points are much less scattered.

Same procedure was followed for the crop type Aman in the region Faridpur and the crop type Boro in the region Khulna. And the scattered-ness shown in the graphs – Figure 4.11e, Figure 4.11f, Figure 4.11i and Figure 4.11j become much less in the graphs - Figure 4.11g, Figure 4.11h, Figure 4.11k and Figure 4.11l.

When the data points were more scattered, Multi-Variable Linear Regression algorithm can not give a better result due to the data points being too scattered and inability to draw a linear line in between them. But KNN algorithm takes the nearest neighbor data points into consideration, so the results of this algorithm were better.

But after we have removed the extremely distant data points from consideration, the Multi-Variable Linear Regression algorithm does give a better result because the data points are now less scattered and a linear line can be drawn in between them. KNN algorithm also gives a little better result because the nearest neighbor number increases and also the nearest neighbors come nearer.

This is the reason why in most cases, Multi-Variable Linear Regression algorithm gives a better result than that of KNN algorithm in the prediction part.

# **Chapter 5**

## **MOBILE APPLICATION**

This chapter takes a brief look at the mobile application we have built. How much feasible our app is, in which sector is it more feasible and which sector it is not, what are the requirements of our app and how much importance do they bear, screenshots of our app and what function are they signifying, features of our app, what limitations did we have while building this – all will be discussed in this chapter.

### **5.1 Feasibility Analysis**

For our project, we did feasibility analysis. This is done in 3 different sectors:

a. **Technological Feasibility:**

The app we are building has all the latest technological support. We made the app in such a way that anyone can use it. Not only high class society people, rural area people with limited technological knowledge will also be able to operate this app. An android platform supported mobile is a must. But apart from that, not much hardware is involved in this system. There are many software components involved in this app. The backend algorithms, the app itself and the information database; latest technological support was provided in all of them. The graphical user interfaces are smooth and to ease up the use in our country for our countrymen, Bangla is the default language of this app.

b. **Economic Feasibility:**

Our app is economically feasible. Why? First of all, we are releasing the alpha version of our app in the play store for free. It might prove economically not so feasible for us. But we believe that in the long run, it will. With more features added, better algorithms to increase accuracy and fulfilling the user requirements that the users demand; we will release better future versions of our app. And they will be suited according to a feasible cost. We did not need to spend much money to collect data for our app, thanks to AIS (Agricultural Information Service), BARC (Bangladesh Agricultural Research Centre), BARI (Bangladesh Agricultural Research Institution), BRRI (Bangladesh Rice Research Institution), BJRI (Bangladesh Jute Research

Institution), SRDI (Soil Resource Development Institute), BBS (Bangladesh Bureau of Statistics) and many other government organizations. They provided us all the necessary data for free. Our group members have previous experience of working in such tasks. And we did not need to add any additional developer/ programmer. So, cost was controlled in this case too.

c. **Organizational Feasibility:**

Our system does not aim to compete with any existing system related to agriculture. There is no such app or even web-supported system in our country so far. Our app is in that case, one of a kind. So, there will be no trend to follow. We will set a trend and a standard for apps of this type. Since it is a new thing, we believe that people will accept it whole-heartedly. The websites related to agriculture and prediction can get benefit from our app, and vice-versa.

## **5.2 Requirement Analysis**

We analyzed the requirements for our app. As we know that requirements are divided in 2 parts – Functional and Non-Functional.

a. **Functional Requirements:**

1. **Smart Phone:** We are building an android mobile app. So, user will need smart phone. More accurately, they will require an android phone to use the app.
2. **Internet Connection:** Our project is a mobile application intended for smart phones. So, user will definitely need data connection to operate the app.
3. **User Authorization:** User authorization is mandatory for this project. Unauthorized access can put the app at risk.
4. **Security:** Overall security for this app is a must. Someone being able to hack this system can create a great problem. So, our motive is to build the application and store data in such a way that even we, ourselves should not be able to hack it.
5. **Database Support:** Our app requires a very good database support. Database being non-functional at any time might create delay in delivering the output or even crash the app. Also, a server-side database with calculations being done regularly in the pre-prediction process is necessary.

b. Non-Functional Requirements:

1. More User-Friendly UI: As this is a prototype, focus on UI was not much. More user-friendly UIs can be designed in the future.
2. Quick Recovery: The app, when crashed; should be able to recover quickly. Taking too much time will get negative impact on user mind. Also, a crash report would give user more reliability.
3. Simultaneous Access: Multiple users may log in to our application in the difference of milliseconds. Our app needs to be ready for high traffic and efficient way of handling them.
4. Faster: The apps' data processing speed should be as faster as possible. Local database (mobile storage) should not need to do too much work. Server side of the app will be given more pressure but that is for a faster running app.
5. Better data support system: The application currently uses basic, free for all database system – MySQL. Getting to use institution-level and more secured database is a need.



### 5.3 User Interfaces

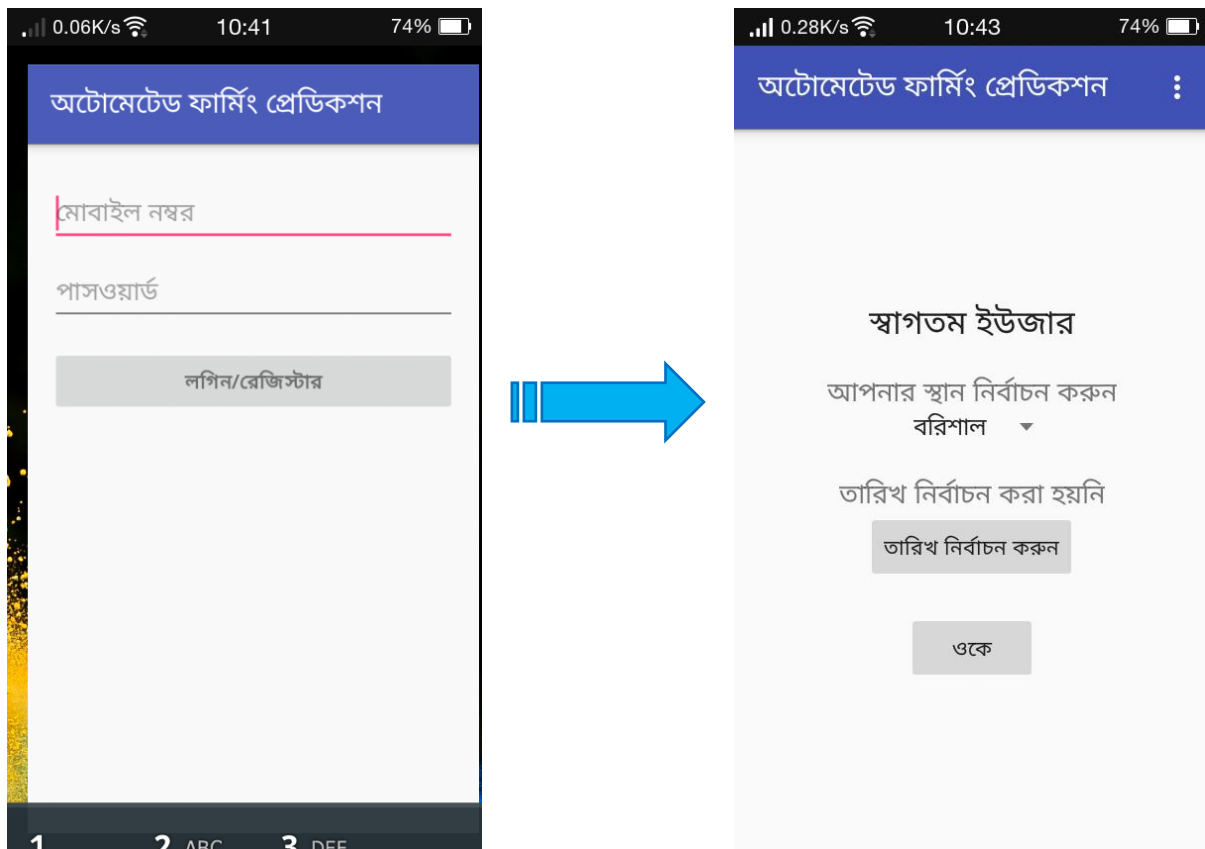


Figure 5.3a: Log In/ Register Figure 5.3b: Welcome UI

Figure 5.3a symbolizes the log in/ register UI of our app. Only once a registered user enters valid mobile number (user name) and password, can s/he be logged in. To register to a new id, one has to input mobile number right, which means one has to give at least 11 numbers. That is the limit we set for mobile number inputs. And user will have to give a password which must be of at least 8 characters. Anything less than 8 characters will be denied as a login password.

Once the user is registered/ logged in, s/he will be directed to the welcome UI of our app (Figure 5.3b). In this UI, user will be provided with 10 options for a region to choose from and a calendar from where the user will choose the probable cropping procedure beginning day. Once user has chosen a valid cropping date, app will now direct the user to a new UI.

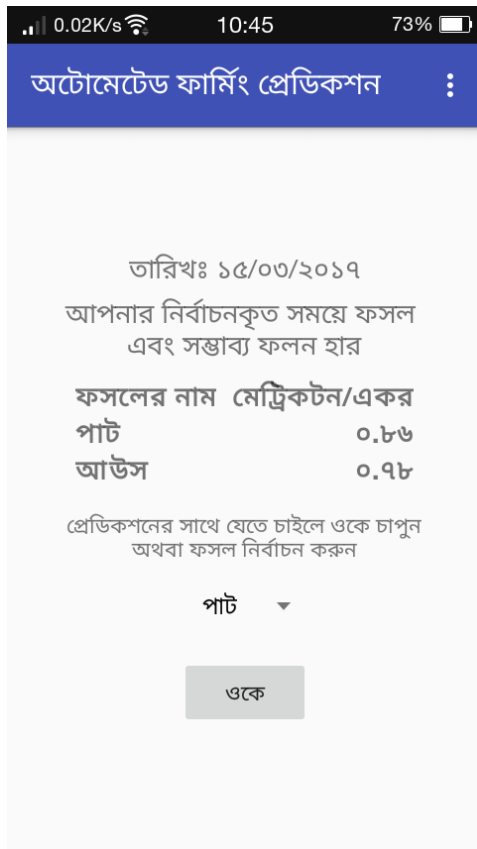


Figure 5.3c: Prediction



Figure 5.3d: Choosing Crop

The user is now at a new UI (Figure 5.3c). In this UI, the app shows possible crop types that can be cultivated in the given time and chosen region by the user. For some cases, there can be multiple crop suggestions. For some cases, there can be only one crop suggestion. Our app suggests the crops in the descending order according to their probable yield rate for next year. The unit of yield rate per unit area is set to metric ton/ acre.

When there is only one crop type possible to cultivate, there will be no option for the user to choose from. But when there are multiple options to choose from, user will have full independence to choose any of the suggested crops (Figure 5.3d). According to the option user has chosen, app will then proceed to the next UI.



Figure 5.3e: Cropping Procedure

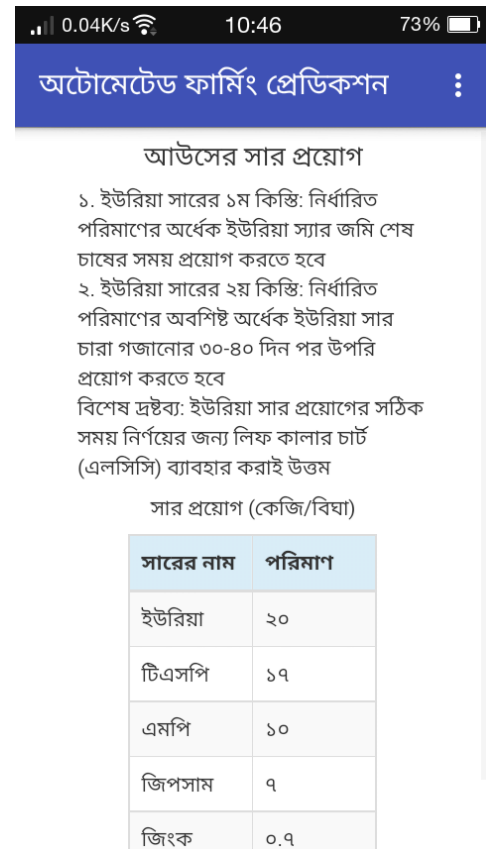


Figure 5.3f: Fertilizer Suggestion

The user can now see the whole cropping procedure of his/her chosen crop type (Figure 5.3e). The whole procedure will be presented in a tabular format. When to sow the seed, when to plant the seedling, when to irrigate and how much, when to apply fertilizer and the amount of fertilizer to be applied, all will be given in this table. The timeline of such procedure is dynamic. How is that represented in the table? The first task – seed sowing day is suggested not as a single day, but as a time period. User can choose any of the dates to start cropping procedure. According to the first task day, the rest will be decided.

User might be interested in the fertilizer amount and application procedure. Or there can be a user who has just begun cropping procedure. There is a button named ‘সার প্রয়োগ বিবরণী’. By clicking this button, user will be directed to a new UI (Figure 5.3f) which will consist of the whole fertilizer application procedure – the component types, their amount, days when they should be applied, how to apply them and lot more. The unit of fertilizer component amount is kg/ bigha (=0.33 acre).



Figure 5.3g: Setting Notification



Figure 5.3h: Reminder of harvesting

From the UI of cropping procedure tabular format (Figure 5.3e), user can click the button ‘সেটনোটিফিকেশন’. Clicking the button will let the user enable the app to set notifications for the important dates, discussed in the cropping procedure UI. There will be a confirmation given to the user (Figure 5.3g). The app will automatically take the date user gave as the input in the Welcome UI (Figure 5.3b), shown clearly in Figure , as the date when user will sow seed/ start the cropping procedure. Taking that date as the beginning date of cropping procedure, app will relate to the date for creating next important dates – irrigation date, fertilizer application date, harvesting date.

In the figure 5.3f, we can see a notification shown in a UI. It gives notification and reminds the user that the harvesting day for his/her chosen crop type – Aus has arrived.



Figure 5.3i: Side Button

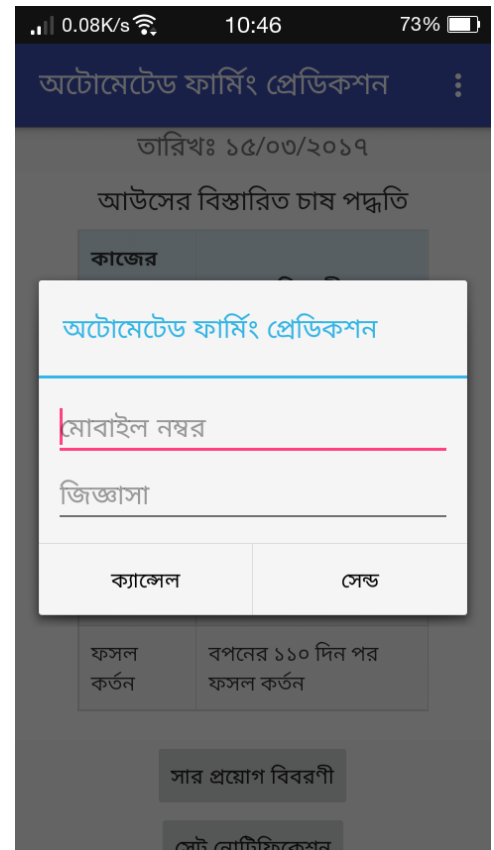


Figure 5.3j: Feedback

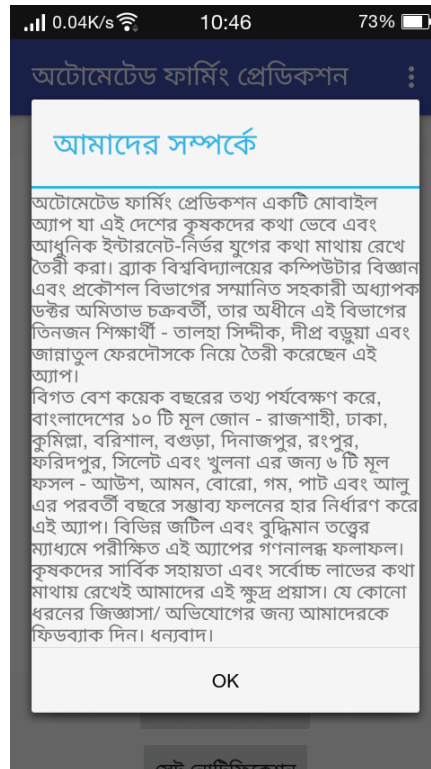
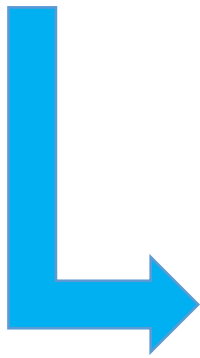


Figure 5.3k: About Section

There is a side-button in our app (Figure 5.3i). This button enables a user to choose from two options – Feedback (ফিডব্যাক) and About (আমাদের সম্পর্কে).

Choosing the feedback option allows a pop-up to occur, with field for mobile number input and the feedback input (Figure 5.3j). User can submit any mobile number; the number does not necessarily have to match with the mobile number used by the user to log in. Once the user submits the review, it will be stored in the database. Depending on whether the review was for the app administrators or the agriculture expert, it will be replied by the admins or forwarded to agricultural experts.

If the user chooses the about section, another pop-up will appear (Figure 5.3k). The pop-up will discuss about the app, its creators, the necessity of the app and a summarized and very much simplified version of its working procedure.

#### Exceptional Cases:

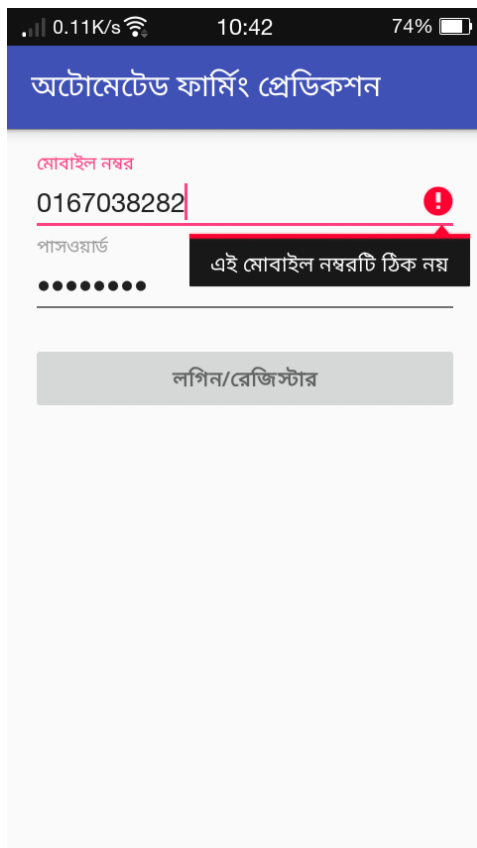


Figure 5.3l: Invalid Username

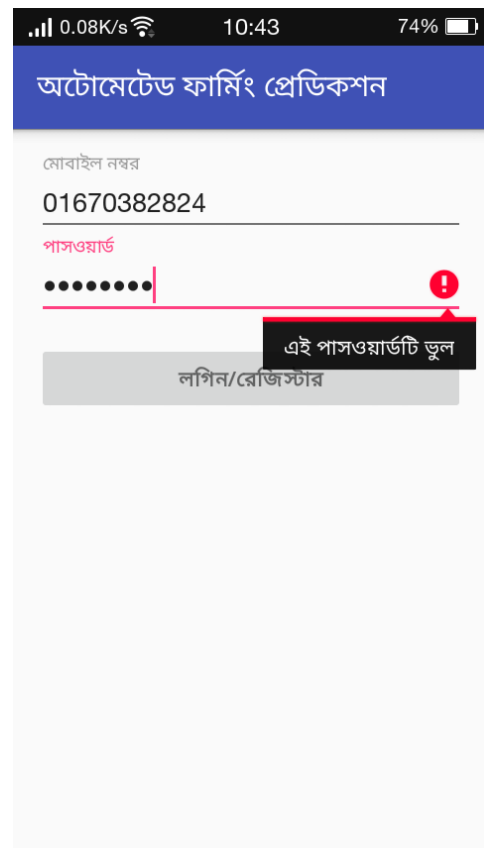


Figure 5.3m: Invalid Password

Figure 5.3l indicates wrong username case. If any user forgets to complete 11 digits for his/her mobile number as the username, app will stop access.

Figure 5.3m shows wrong password case. If a registered user gives wrong password, access will be stopped. If a new user gives a password of less than 8 characters, access will be stopped.

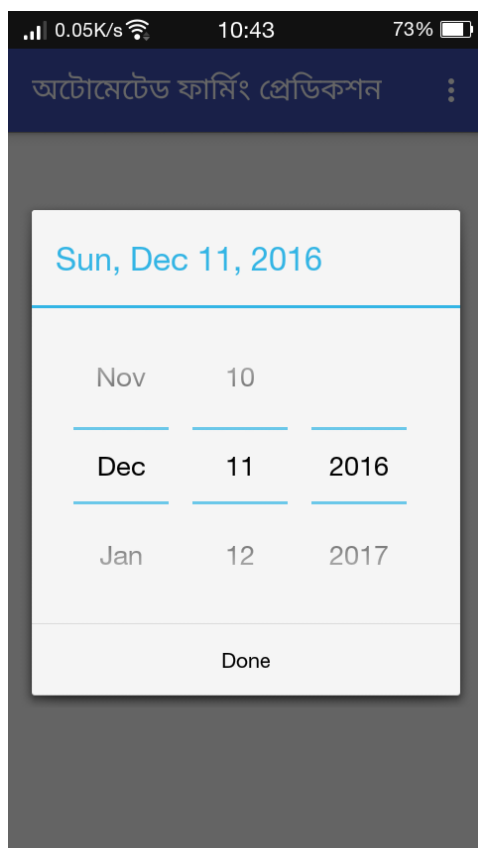


Figure 5.3n: Date Choosing

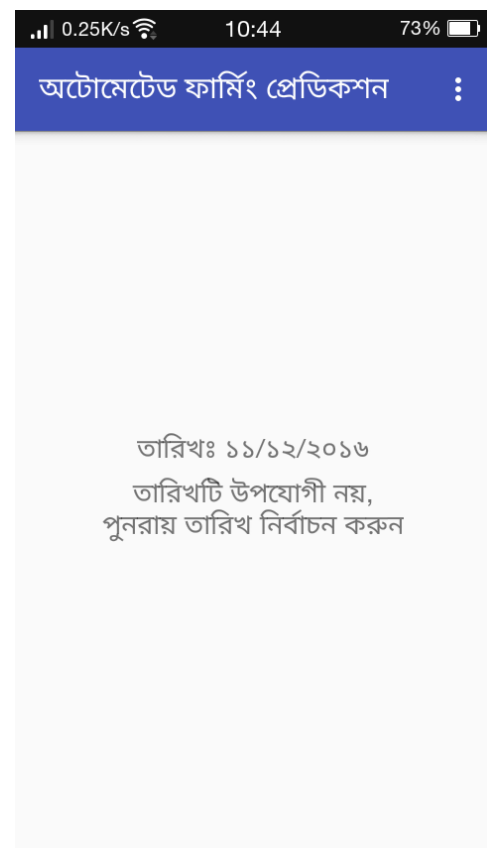


Figure 5.3o: Wrong Date

If a user gives a date which is not included in cropping time-period of any of the 6 suggested crops (Figure 5.3n), app will give an error output in a different UI. It will say in Bengali, “The date inputted is not eligible for cultivating, please select a date again.”

## 5.4 Features

1. One-Click Initiation: Our app allows user to register and login at the same time. User will not need to register and then go back, and then login.
2. Takes Date and Location Input – Our app takes the current location of the user and the expected cropping beginning date for the certain crop.
3. Machine Learning Algorithm – Based on the input given, app decides which crops can be processed. And then it runs a machine learning algorithm – Multiple Linear Regression.
4. Prediction: By running the machine learning algorithm in the backend, it predicts possible yield rate in metric ton per acre for each crop.
5. Freedom to Choose: App then allows the user to choose one of the multiple crops (if available).
6. Farming Procedure in a Nutshell: After user chooses a certain crop, app then shows detailed planning of the farming procedure for the chosen crop.
7. Fertilizer Suggestion: There's an option to see fertilizer suggestion throughout the farming period of the chosen crop. Choosing it, user enables the app to show them fertilizer applying procedure throughout the farming period.
8. Creates Schedule: With permission from the user, app creates reminders or notification systems for important dates. For example, the day of sowing seed, day of fertilizer application.
9. Notification: On the important days, app will notify the user through a reminder that s/he has a task to do.
10. Feedback: There is an option for user to give feedback. Anything wrong about the app or its output or activities can be reported by user.
11. Uniqueness: As this project and it's idea is the first ever in the concept of our country, the app has its own uniqueness.
12. Large Database: Our app has a database with data of 15 years in the backend to make more accurate predictions.
13. Internal Server: The app runs the machine learning algorithm in the backend and generates prediction for the future. Then it stores the data in a local server.
14. Less Local Processing: We tried to minimize the involvement of local database and processing functions, so that the app could be faster and run smoothly.



15. Android App: Our project is a mobile application built in specifically for android platform.
16. Bangla: We kept in mind, the ease of local users. That is why the operating interface language of our app is Bangla.

## **5.5 Limitations**

1. Prototype: We are presenting a prototype version of a large, grand and broad idea. More time and effort could make it a better deliverable. We have no standard or model to follow. In a way, we shall create a standard to be followed in the future!
2. Data size: The data size we have collected is not quite large. The data size is of 15 years (starting from 2000 to 2014). A prediction project like ours need a huge amount of data in the backend to make the prediction more accurate. We feel that we require data of 10 to 15 more years to declare this app almost fully authentic.
3. Analog way of data storage: The data size could have still been larger. But data of years before 2008 are not digitalized. We had to run to many places, search many libraries, read many books to find out data from 2000 to 2007. And the data before them are too scattered that we could not collect. If these data were stored digitally, we could have presented a more accurate predicting system.
4. Less relevant knowledge: We the researchers did not have much knowledge of mobile application building or agriculture beforehand. We had to learn about these and then get to work.
5. Lack of real future data: If we had access to weather data of last some months of 2016 and 2017, we could generate prediction of the crops for 2017. We could generate prediction for future. But these data are really rare and in the context of our country, these data are too difficult to find. We did not have enough funding backup to buy weather data as well.

## Chapter 6

### CONCLUSION AND FUTURE WORKS

In this chapter, we will look at some possible modifications, future scope of our application. What we missed, what more could we add, how could this app be of more help for serving the farmers of our country; we will see here. And finally, we will summarize our work and the paper with a conclusion.

#### 6.1 Future Works

1. Multi-Platform app: Our project is currently an android-platform mobile app. It can be rebuilt in a multi-platform structure so that windows and iPhone users can also use this app.
2. Image Processing: Reason behind us not including pesticide and disease remedies is the uncertainty of these variables and their behaviors. A crop can suffer different diseases and be attacked by different type of insects. To know the actual problem, image processing can be incorporated to this project.
3. Data Size: After 10 years, the data size will be of 25 years' minimum. And with a data size that large, the app will be able to give more accurate predictions.
4. Improving the sample area: Currently, our app works for 10 major regions. In future, all of the 64 districts can be considered as the sample area. In that case, the prediction will be more accurate for the region/ district considered.
5. Crop Specialization: Our app now considers cropping of 6 different crop types. There are hundreds more crop types and so many sub-species under each of them. By considering the crop types and their sub-species, the app can give prediction for more crop types.
6. Resources: Our initial plan was to track the location where resources like fertilizer, pesticides and such can be found. Though we could not do this, it would be more cost effective if the location of such resources, at a cheaper price can be found through this app.
7. Video Demo: Someone with no prior knowledge might be interested in farming. For those people, video demo of different procedures like irrigation, fertilizer applying

and such can be included in the system. Or a link directing to the videos already stored in some other websites can also be feasible idea.

8. Simultaneous Access: The app is in a prototype version. So, it might not be able to handle many users at the same time. Handling more simultaneous users through improving app performance is a future scope.
9. More Security: We are currently using basic security protocols for the app and the database support. But when the app is considered for real life implementation, 256-bit aes encrypted security system can be considered.
10. Offline Mode: The app requires data connection to be operated. An offline mode of the mobile app can be implemented in the future.
11. Location Tracker: By using gps and such technology, user will no longer be required to give their location as input. Location of the user will be tracked immediately as the user logs in to the app.
12. Web-Version: A web-version of this app is a need of the hour. Many people prefer browsing through desktop/ laptop. For them, a web-version of the app can be implemented.
13. Live conversation: Currently we are taking feedback from the users in old fashioned way. Live conversation option can be incorporated to the app, so that user can contact agricultural extension people anytime as they need help with anything.

## **6.2 Conclusion**

We are building a prototype of what can be a great innovation for the future. Being dependent on agriculture for a long time, our country has not seen much collaboration between technology and agriculture so far. There are some websites and also a few mobile applications already in use, for agriculture in this country. But we aim at a future when almost everyone uses a smart mobile phone. We intend to make cropping prediction and procedures digital. Agriculture Extension also welcomes this initiative and they think that this will be a standard to be followed in the future too. Our step is very little but we hope that this is the beginning to something big.

## Reference

- [1] Yearbook of Agricultural Statistics - (2008-2015) [Online]. <http://bbs.gov.bd> (Accessed: 12 December 2016).
- [2] Shou-chuan Yang, "A Search Algorithm and Data Structure for an Efficient Information System," University of Wisconsin Madison, Wisconsin.
- [3] Sunil Ray. (2015) Analytics Vidhya. [Online]. available at: <https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/> (Accessed: 12 December 2016).
- [4] Trading Economics. (1994-2016) Trading Economics. [Online]. Available at: <http://www.tradingeconomics.com/bangladesh/gdp-growth> (Accessed: 15 August 2016).
- [5] IRRI. (2012) IRRI. [Online] available at: <http://irri.org/news/media-releases/feeding-rice-just-got-easier-with-smartphones> (Accessed: 12 December 2016).
- [6] Algorithm (Multiple Linear Regression).[Online]. Available at: <http://www.originlab.com/doc/Origin-Help/Multi-Regression-Algorithm> (Accessed: 12 December 2016).
- [7] K-Nearest Neighbors Algorithm [Online]. Available at: [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm#k-NN\\_regression](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#k-NN_regression) (Accessed: 12 December 2016).
- [8] Farhad Soleimani Gharehchopogh, Tahmineh Haddadi Bonab, Seyyed Reza Khaze, "A Linear Regression Approach to Prediction of Stock Market Trading Volume: A Case Study," International Journal of Managing Value and Supply Chains (IJMVSC), Vol. 4, No. 3, September 2013, DOI: 10.5121/ijmvsc.213.4303.
- [9] Lucas Nunno, "Stock Market Price Prediction Using Linear and Polynomial Regression Models," University of New Mexico Computer Science Department Albuquerque, New Mexico, United States [[lnunno@cs.unm.edu](mailto:lnunno@cs.unm.edu)].
- [10] Zizhen Yao, Walter L Ruzzo, "A Regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data," BioMed Central Ltd. 2006

[11] Matti Maltamo, Annika Kangas, “Methods based on  $k$ -nearest neighbor regression in the prediction of basal area diameter distribution,” Canadian Journal of Forest Research, 1998